**Jari Råman**

# Regulating Secure Software Development

## Analysing the potential regulatory solutions for the lack of security in software

Academic Dissertation to be presented,
with the permission of the Faculty of Law of the University of Lapland,
for public discussion in Auditorium 2, Yliopistonkatu 8, Rovaniemi,
on May 26th, 2006, at 12 o´clock.

University of Lapland
Faculty of Law



Distributor: Lapland University Press
P.O. Box 8123
FI-96101 Rovaniemi

tel. + 358 16 341 2924, fax + 358 16 341 2933
julkaisu@ulapland.fi
www.ulapland.fi/publications

Paperback
ISBN 952-484-034-0
ISSN 0788-7604

PDF
ISBN 952-484-053-7
ISSN 1796-6310
www.ulapland.fi/unipub

*"…if a person can't feel safe, he can never be free…"*
*(ADA Richard Bay in the drama series "The Practice",*
*Season 4: Episode 14 – "Checkmates")*

# Preface

This is what an assistant district attorney (ADA) told to a colleague about the role of prosecutors in order to cheer her up after losing another case to a devious defence counsel. Who said that watching courtroom drama, especially American series, is a waste of time; that they do not teach anything about lawyers' work in continental law countries? Well... Maybe they do not, but this prosecutor certainly understood a part of the essence of freedom and safety and I got to spend amusing moments in front of a TV.

This ADA was talking about personal safety, but the same statement is true also for security in general. In the era of Internet and the discussion of its inherent freedom, this really is a crucial statement. If we want to preserve at least some of the former imagined liberty of the Internet its time to take its security seriously.

But it is also said that they that can give up essential liberty to obtain a little temporary safety deserve neither liberty nor safety. At first look these arguments seem contradictory. In fact, they are not. It is true that if you live in constant fear, you will not be able to enjoy the freedom you have. Likewise, if you give a way part of your liberty to obtain safety, you will end up with neither of them. The issue is about balance. These arguments are certainly true in the tangible world where we live in, but they are true also for the networked society and the virtual worlds.

This is a legal scientist journey into security in the networked world. This really is a *journey*. One which I originally took with another destination in mind. To be honest, *this is not the thesis I intended to make*. It started out as a study of the role of information security in the central tenets of the constitutional state and the system of basic rights together with the place of information

security among central legal principles. If anything, in western constitutional democracies we build liberty by setting society upon a certain constitution to deal with the balancing. As I proceeded, I started doubting the usefulness of such an approach to other than lawyers in constitutional states like the Finnish; to also whom it is only of limited theoretical interest.

While trying to understand to role of the constitutional rights arguments I faced the role of the coherence argument of law and noticed the line between the internal and external perspectives on law. Then something peculiar occurred; the law seemed to hamper the information security research and practice on many occasions, the one it was supposed to be enhancing.

At the same time another interesting problem occurred. The 'law' and the legislator seemed to assume that the underlaying infrastructure and the components used therein are secure. Many of the legal provisions on transactions, contracting and on the use of constitutional rights in the networks in general seemed to build on this assumption. At the same time the infrastructure of the network society was widely recognised to be insecure by the information security community.

This puzzled me so much that I had to go deeper. I wanted to understand why. Being a young and inexperienced researcher I threw myself headlong on the issue. What started out as a research at the very core of legal scholarship, legal and especially constitutional theory, turned out to study the fringes of the law.

This is the outcome of that journey; an odyssey one might say. So much for my discipline... (Un)fortunately, this still is somewhat visible in the text.

Before I can rejoin the original path with the wisdom, especially of the law, gathered during the many encounters on my journey of exploration, or make another one, the time has come to give thanks.

My supervisor, Professor *Ahti Saarenpää*, deserves my gratitude above all for the academic freedom and encouragement. The high scholarly example you set forced me to do my best. Dissertation examiners Professor *Kauko Wikström* from the

University of Turku and Professor *Gerald Quirchmayr* from the University of Vienna made valuable comments to draft versions. I have taken most of them into account. Professor Quirchmayr kindly agreed to act as the academic opponent at the public defence.

The odyssey would not have been the same if I had not come in contact with information security researchers from the Department of Information Processing Science at the University of Oulu and from the Oulu University Secure Programming Group. The former gave me an understanding of what information security is really all about. The latter not only acquainted me with the life of a bug, but also gave me a glimpse of what academic group work can be at its best.

The research could not have been possible without financial support. The project *Scarcity of Justice* funded by the Academy of Finland gave me the original possibility to learn to know myself as a researcher and to make the initial wonderings in the dark. I am obliged to the Institute for Law and Informatics at the Faculty of Law. A thankyou goes also to the Rector of the University of Lapland and the Finnish Lawyers' Association. At the final stages the Faculty of Law gave me a position as an assistant in legal informatics. I was able to finalise my thesis without overly burdensome administrative tasks largely due to the understanding of the acting professor *Rauno Korhonen*.

Friends have really made the journey. A warm thankyou goes to *Annamari*, *Anu* and *Pekka* for just being yourselves. Thanks also to my parents *Tuula* and *Kyösti*, to my siblings *Hanna*, *Henri* and *Petri*, and to friends, relatives and colleagues not especially mentioned.

Words are not enough for my nearest and dearest *Mervi, Sofia* and *Selina*.

Rovaniemi, May 2006

*Jari Råman*

## Contents

# Sources

## Literature

Aarnio A (1989) *Laintulkinnan teoria. Yleisen oikeustieteen oppikirja*, Werner Söderström Oy, Juva

Aarnio A (1983) Some Conceptual Foundations of Legal Policy Research, in *Philosophical Perspectives in Jurisprudence*, Acta Philosophica Fennica Vol. 36, Philosophical Society of Finland, Helsinki, p. 222-238

Abrahamsson P (2002) *The Role of Commitment in Software Process Improvement*, Oulu University Press, Oulu

Abrahamsson P, Salo O, Ronkainen J and Warsta J (2002) *Agile Software Development Methods: Review and Analysis*, VTT Publications 478, Technical Research Centre of Finland, Espoo
http://www.inf.vtt.fi/pdf/publications/2002/P478.pdf [23.2.2006]

Acquisti A and Grossklags J (2003) *Losses, Gains, and Hyperbolic Discounting: An Experimental Approach to Information Security Attitudes and Behaviors*, paper presented at the Second Annual Workshop on Economics and Information Security, Robert H. Smith School of Business, University of Maryland, May 29-30, 2003, available at
http://www.heinz.cmu.edu/~acquisti/papers/acquisti_grossklags_eis_refs.pdf [21.2.2006] published in Camp J and Lewis S (eds., 2004) *The Economics of Information Security* (Advances in Information Security), Kluwer

Adler MD (2000) Beyond Efficiency and Procedure: A Welfarist Theory of Regulation, *Florida State University Law Review*, 28(1): 241-339

Adler MD (2000) Expressive Theories of Law: A Skeptical Overview, *University of Pennsylvania Law Review*, 148(5): 1363-1502

Adler MD (2000) Linguistic Meaning, Nonlinguistic 'Expression' and the Multiple Variants of Expressivism: A Reply to Professors Anderson and Pildes, *University of Pennsylvania Law Review*, 148(5): 1577-1595

Ahonen P, Eronen J, Holappa J, Kajava J, Kaksonen T, Karjalainen K, Karppinen K, Rapeli M, Röning J, Sademies A, Savola R, Uusitalo I and Wiander T (2005) *Information Security Threats and Solutions in the Mobile World. The Service Developer's Perspective*, VTT Research Notes 2308, VTT Technical Research Centre of Finland, Espoo. Available also at the web pages of the Development Programme on Trust and Information Security in Electronic Services (LUOTI) of the Finnish Ministry of Transport and Communications, at http://www.luoti.fi/publish.html [22.12.2005]

Akerlof GA (1970) The Market for "Lemons": Quality Uncertainty and the Market Mechanism, *Quarterly Journal of Economics*, 84(3): 488-500

Anderson ES and Pildes RH (2000) Expressive Theories of Law: A General Restatement, *University of Pennsylvania Law Review*, 148(5): 1503-1576

Anderson R (2003) *Cryptology and Competition Policy-Issues with 'Trusted Computing'*, Paper presented at the 2nd Annual Workshop on Economics and Information Security, University of Maryland, May 29-30

Anderson R (2001) *Why Information Security is Hard – An Economic Perspective*, paper presented at 17th Annual Computer Security Applications Conference, December 10-14, New Orleans, Louisiana

Antunes G and Hunt AL (1980) The Impact of Certainty and Severity of Punishment, in Evan WM (ed.) *The Sociology of Law. A Social-Structural Perspective*, The Free Press, New York, p. 185-197. Excerpt from Antunes G and Hunt AL (1973) The Impact of Certainty and Severity of Punishment on Levels of Crime in American States: An Extended Analysis, *Journal of Criminal Law and Criminology*, 64: 486-493

Arbaugh B (2002) Security: Technical, Social, and Legal Challenges, *IEEE Computer*, 35(2): 109-111

Arbaugh WA, Fithen WL and McHugh J (2001) Windows of Vulnerability: A Case Study Analysis, *IEEE Computer*, 33(12): 52-59, also available at http://www.cs.umd.edu/~waa/pubs/Windows_of_Vulnerability.pdf [23.2.2006]

Arhippainen L (2003) *Use and integration of third-party components in software development*, VTT Publications 489, Technical Research Centre of Finland, Espoo, http://www.vtt.fi/inf/pdf/publications/2003/P489.pdf [16.2.2006]

Arora A, Krishnan R, Nandkumar A, Telang R and Yang Y (2004) Impact of Vulnerability Disclosure and Patch Availability — An Empirical Analysis, paper presented at *The Third Annual Workshop on Economics and Information Security* (WEIS04), May 13-14, 2004, University of Minnesota, Minneapolis, USA,
http://www.dtc.umn.edu/weis2004/telang.pdf [23.2.2006]

Ashish A, Caulkins JP and Telang R (2003) *Sell First Fix Later: Impact of Patching on Software Quality*, Carnegie Mellon University Working Paper, January, http://www.heinz.cmu.edu/~rtelang/patchingF.pdf [23.2.2006]

Aubert V (1976) *Rettens sosiale funksjon*, Universitetsforlaget, Oslo

Aubert V (1966) Some Social Functions of Legislation, in Aubert V (ed. 1969) *Sociology of Law. Selected Readings*, Penguin Books, Middlesex (abridged)

Ayres I and Braithwaite J (1992) *Responsive Regulation: Transcending the Deregulation Debate*, Oxford University Press, Oxford

Baldwin R (1997) Regulation: After Command and Control, in Hawkins K (ed.) *The Human Face of Law*,Oxford University Press, Oxford

Baldwin R (1995) *Rules and Government*, Clarendon Press, Oxford

Baldwin R and Cave M (1999) *Understanding Regulation. Theory, Strategy, and Practice*, Oxford University Press, New York

Baldwin R, Scott C and Hood C (eds., 1998) *A Reader on Regulation*, Oxford: Oxford University Press

von Bar C and Drobning U (2001) *Study on Property Law and Non-contractual Liability Law as they relate to Contract Law*, Submitted to the European Commission, Health and Consumer Protection DG, SANCO B5-1000/02/000574, http://europa.eu.int/comm/consumers/cons_int/safe_shop/fair_bus_pract/cont_law/study.pdf [23.2.2006]

Bar-Gill O and Fershtman C (2004) Law and Preferences, *Journal of Law, Economics and Organisation*, 20(2): 331-352

Baron DP (2001) Private Politics, Corporate Social Responsibility, and Integrated Strategy, *Journal of Economics & Management Strategy*, 10(1): 7-45

Baron DP (2003) Private Politics, *Journal of Economics & Management Strategy*, 12(1): 31-66

Baskerville R (1992) The Developmental Duality of Information Systems Security, *Journal of Management Systems*, 4(1): 1-12

Baskerville R (1993) Information Systems Security Design Methods: Implications for Information Systems Development, *ACM Computing Surveys*, 25(4): 375-414

Baskerville R, Levine L, Pries-Heje J, Ramesh B and Slaughter S (2001) How Internet Software Companies Negotiate Quality, *IEEE Computer*, 34(5): 51-58

Baskerville R and Pries-Heje J (2001) Racing the e-bomb: How the Internet is Redifining Information Systems Development Methodology, in FitzGerald B, Russo N and DeGross J (eds.), *Realigning Research and Practice in IS Development: The Social and Organisational Perspective*, Kluwer, New York, p. 49-68

Beales H, Craswell R, and Salop SC (1981) The Efficient Regulation of Consumer Information, *Journal of Law & Economics*, XXIV(3): 491-539, reprinted in Ogus AI (ed., 2001) *Regulation, Economics, and the Law*, International Libraty of Critical Writings in Economics Series, No. 137, Edward Elgar Publishing, UK, p. 160-209

Bemelmans-Videc M-L (1998) Introduction: Policy Instruments Choice and Evaluation, in Bemelmans-Videc M-L, Rist RC and Vedung E (eds.) *Carrots, Sticks & Sermons. Policy Instruments & Their Evaluation*, Transaction Publishers, New Brunswick, p. 1-21

Bemelmans-Videc M-L and Vedung E (1998) Conclusions: Policy Instruments Types, Packages, Choices, and Evaluation, in Bemelmans-Videc M-L, Rist RC andVedung E (eds.) *Carrots, Sticks & Sermons. Policy Instruments & Their Evaluation*, Transaction Publishers, New Brunswick, p. 249-275

Bender D (1991) Computer Software Products Liability – The United States Perspective, in Meijboom AP and Prins C (eds.) *The Law of Information Technology in Europe 1992*, Kluwer, Deventer, p. 207-225

Berman PS (2000) Cyberspace and the State-Action Debate: The Cultural Value of Applying Constitutional Norms to 'Private' Regulation, *University of Connecticut School of Law Working Paper Series*, Working Paper 9 (June 1st), available at http://lsr.nellco.org/uconn/ucwps/papers/9 [updated 5.12.2005, visited 23.2.2006]. Published also at *University of Colorado Law Review*, 71(4): 1263-1310

Better Regulation Task Force (2005) *Routes to Better Regulation: A Guide to Alternatives to Classic Regulation*, Better Regulation Commission, London, available at http://www.brc.gov.uk/publications/routes.asp [9.3.2006]

Better Regulation Task Force (2005) *Get Connected: Effective Engagement in the EU*, Better Regulation Commission, London, available at http://www.brc.gov.uk/publications/getconnectedentry.asp[9.3.2006]

Black HC (1998) *Black's Law Dictionary*, 6th edition, 13th reprint, West Publishing Co., St. Paul

Black J (1996) Constitutionalising Self-Regulation, *Modern Law Review*, 59(1): 24-56

Black J (2002) Critical Reflections on Regulation, *CARR (Centre for Analysis of Risk and Regulation) Discussion Paper* no: 4, January, London School of Economics and Political Science, London, http://www.lse.ac.uk/collections/CARR/pdf/Disspaper4.pdf [16.2.2006]

Black J (2001) Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a 'Post-Regulatory' World, *Current Legal Problems*, 54: 103-146

Black J (2003) Enrolling Actors in Regulatory Systems: Examples from UK Financial Services Regulation, *Public Law*, Spring, Sweet & Maxwell, London, p. 63-92

Black J (2004) Law and Regulation: The Case of Finance, in Parker C, Scott C, Lacey N and Braithwaite J (eds.) *Regulating Law*, Oxford University Press, Oxford, p. 33-60

Black J (2003) Mapping the Contours of Contemporary Financial Services Regulation, *CARR (Centre for Analysis of Risk and Regulation) Discussion Paper* No. 17, London School of Economics and Political Science, London, http://www.lse.ac.uk/collections/CARR/pdf/Disspaper17.pdf [16.2.2006]

Black J (2000) Proceduralising Regulation: Part I, *Oxford Journal of Legal Studies*, 20(4): 597-614

Black J (2001) Proceduralizing Regulation: Part II, *Oxford Journal of Legal Studies*, 21(1): 33-58

Blackburn JD, Scudder GD and Wassenhove LN (1996) Improving Speed and Productivity of Software Development: A Global Survey of Software Developers, *IEEE Transactions on Software Engineering*, 22(12): 875-886

Boehm BW (1981) *Software Engineering Economics*, Prentice-Hall, New Jersey

Boehm BW and Sullivan KJ (2000) Software Economics: A Roadmap, in Finkelstein A (ed.) *The Future of Software Engineering, 22nd International Conference on Software Engineering*, ACM, New York, June

Botting RJ (1997) On the Economics of Mass-Marketed Software, *Proceedings of the 19th International Conference on Software Engineering* (ICSE), May 17-23, Boston, Massachusetts, p. 465-471

Boyle J (1997) *Foucault in Cyberspace: Surveillance, Sovereignty and Hard-Wired Censors*, available at http://www.law.duke.edu/boylesite/foucault.htm [10.19.2005], published also in *University of Cincinnati Law Review*, 66: 177-205

Bradgate R (1999) Beyond the Millennium – The Legal Issues: Sale of Goods Issues and the Millennium Bug, *Journal of Information, Law and Technology* (JILT), 1999(2), available at http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/1999_2/bradgate/ [27.3.2006]

Brady RM, Anderson RJ, and Ball RC (1999) *Murphy's law, the fitness of evolving species, and the limits of software reliability*, University of Cambridge, Computer Laboratory, Technical report No. 471, September, available at http://www.cl.cam.ac.uk/TechReports/UCAM-CL-TR-471.pdf [21.2.2006]

Breyer SG (1982) *Regulation and its Reform*, Harvard University Press, Cambridge

Brownlee N and Guttman E (1998) *Expectations of Computer Security Incident Response*, Internet Engineering Task Force (IETF), Request for Comment (RFC) 2350, http://www.ietf.org/rfc/rfc2350.txt [23.1.2006]

Brüggemeier G (1997) The Control of Corporate Conduct and Reduction of Uncertainty by Tort Law, in Baldwin R (ed.) *Law and Uncertainty. Risks and Legal Processes*, Kluwer Law International, The Hague, p. 57-75

Bryde Andersen M (1988) *Edb og ansvar*, Jurist- og Økonomforbundets forlag, København

Bryde Andersen M (2005) *IT-retten*, 2nd ed., Gjellerup, København. The 1st edition is also available at http://www.it-retten.dk [4.1.2006]

Burnett R (2005) Legal Risk Management for the IT Industry, *Computer Law & Security Report*, 21(1): 61-67

Burrel G and Morgan G (1979) *Sociological Paradigms and Organisational Analysis. Elements of the Sociology of Corporate Life*, Heinemann, London

Burrows P (1994) Products Liability and the Control of Product Risk in the European Community, *Oxford Review of Economic Policy*, 10(1): 68-83

Calderini M, Cantamessa M and Palmigiano A (2003) *Analysis of the Economic Impact of the Development Risk Clause as provided by Directive 85/374/EEC on Liability for Defective Products*, Study for the European Commission, Contract No. ETD/2002/B5, http://europa.eu.int/comm/enterprise/regulation/goods/docs/liability/2004-06-dev-risk-clause-study_en.pdf [23.2.2006]

Callaghan D and O'Sullivan C (2005) Who Should Bear the Cost of Software Bugs? *Computer Law & Security Report*, 21(1): 56-60

Cambell K, Gordon LA, Loeb MP, and Zhou L (2003) The Economic Cost of Publicly Announced Information Security Breaches: Empirical Evidence From the Stock Market, *Journal of Computer Security*, 11(3): 431-448

Camp JL and Wolfram C (2000) *Pricing Security*, Proceedings of the CERT Information Survivability Workshop, Boston, MA Oct. 24-26, 2000, p. 31-39, http://www.ljean.com/files/isw.pdf [23.2.2006]

Carlshamre P (2001) *A Usability Perspective on Requirements Engineering. From Methodology to Product Development*, Linköping Studies in Science and Technology nr 726, Linköping University, Linköping, available at http://www.diva-portal.org/liu/theses/abstract.xsql?dbid=4976 [16.2.2006]

Carlshamre P and Regnell B (2000) Requirements Lifecycle Management and Release Planning in Market-Driven Requirements Engineering Processes, in Tjoa AM, Wagner RR, and Al-Zobaidie A (eds.) *Proceedings of the 11th International Workshop on Database and Expert Systems Applications Process*, IEEE Computer Society Press, Los Alamitos, CA

Carmel E and Sawyer S (1998) Packaged software development teams: What makes them different?, *Information Technology and People*, 11(1): 7–19

Castrén M (1997) EU-Suomen markkinaoikeus, Kauppakaari, Helsinki

Cavusoglu H, Cavusoglu H and Raghunathan S (2005) Emerging Issues in Responsible Vulnerability Disclosure, paper presented in the *Fourth Workshop on the Economics of Information Security*, Kennedy School of Government, Harvard University, 2-3 June 2005, available at http://infosecon.net/workshop/pdf/cavusoglu.pdf [10.2.2006]

Chandler D (1995) *Technological or Media Determinism*, part "Reification", The Media and Communications Studies Site, hosted by the University of Wales, Aberystwyth, http://www.aber.ac.uk/media/Documents/tecdet/tdet05.html (last modified 11.4.2000, visited 10.10.2005)

Chandler JA (2005) *Improving Software Security: A Discussion of Liability for Unreasonably Insecure Software*, Securing Privacy in the Internet Age, Stanford University Press, 2005, at http://ssrn.com/abstract=610041 [21.2.2006]

Chandler JA (2004) Security in Cyberspace: Combatting Distributed Denial of Service Attacks, *University of Ottawa Law & Technology Journal*, 1: 231-261, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=596667 [10.2.2006]

Coleman JL (1988) *Markets, Morals and the Law*, Cambridge University Press, Cambridge

Coleman JL (1991) Rules and social facts, *Harvard Journal of Law & Public Policy*, 14(3): 703- 726

Coleman JL and Leiter B (1996) Legal Positivism, in Patterson D (ed.) *A Companion to Philosophy of Law and Legal Theory*, Blackwell Publishers, Oxford, p. 241-261

Collins H (2004) Regulating Contract Law, in Parker C, Scott C, Lacey N and Braithwaite J (eds.) *Regulating Law*, Oxford University Press, Oxford, p. 13-33

Collins H (1999) *Regulating Contracts*, Oxford University Press, New York

Connolly DR (1994) Insurance: The Liability Messenger, in Hunziker JR and Jones TO (eds.) *Product Liability and Innovation: Managing Risk in an Uncertain Environment*, Washington DC, National Academy Press, p. 131-137 http://www.nap.edu/books/0309051304/html/ [23.2.2006]

Cooter R (2001) Do good laws make good citizens? An economic analysis of internalizing legal values, *Independent Institute Working Paper No. 33*, available at
http://www.independent.org/publications/working_papers/
article.asp?id=744 [updated and visited 23.2.2006]

Cooter R (1995) Law and Unified Social Theory, *Journal of Law and Society*, 22(1): 50-67

Cooter R and Ulen T (2000) *Law and Economics*, 3rd ed., Addison-Wesley, Reading, Massachusett

Cusumano MA and Shelby R (1995) *Microsoft Secrets*, The Free Press, New York

Cusumano MA and Yoffie DB (1998) *Competing on Internet Time: Lessons from Netscape and Its Battle with Microsoft*, The Free Press, New York

Dahlstedt ÅG, Karlsson L, Persson A, Natt och Dag J and Regnell B (2003) Market-Driven Requirements Engineering Processes for Software Products – a Report on Current Practices, paper presented in International Workshop on COTS and Product Software: Why Requirements Are So Important (RECOTS), 10 September 2003, held in conjunction with the 11th IEEE International Requirements Engineering Conference, September 8-12, 2003, Monterey Bay, California, USA, available at
http://www-lsi.upc.es/events/recots/papers/Dahlstedt.pdf [21.2.2006]

Daintith T (1987) Law as Policy Instrument: A Comparative Perspective, in Daintith T (ed.) *Law as an Instrument of Economic Policy: Comparative and Critical Approaches*, European University, Series A, No. 7, Walter de Gruyter, Berlin, p. 3-56

Daughety AF and Reinganum JF (1995) Product safety: liability, R&D and signaling, *The American Economic Review*, 85(5): 1187-1206

Daughety AF and Reinganum JF (2005) Secrecy and Safety, *The American Economic Review*, 95(4): 1074-1091. Also available at http://www.vanderbilt.edu/Econ/faculty/Daughety/DaughetyandReinganumSecrecyandSafetyJan2005.pdf [29.3.2006]

DeLeon P (1999) The Stages Approach to the Policy Process:What Has It Done? Where Is It Going?, in Sabatier PA (ed.) *Theories of the Policy Process*, Westview Press, Oxford, p. 19-35

DeLong B and Froomkin M (2000) Speculative Microeconomics for Tomorrow's Economy, *First Monday*, 5(2), http://www.firstmonday.org/issues/issue5_2/delong/index.html [22.2.2006]

Devanbu P and Stubblebine S (2000) Software Engineering for Security: a Roadmap, in Finkelstein, A (ed.) *The Future of Software Engineering, 22nd International Conference on Software Engineering*, ACM, New York, June, p. 225-241

Dharmapala D and McAdams RH (2001) The Condorcet Jury Theorem and the Expressive Function of the Law: A Theory of Informative Law, *University of Illinois College of Law, Law and Economics Working Paper Series*, Working Paper No. 00-19, February, p. 1, available at http://www.law.uiuc.edu/publications/ssrn/articles/00-19%20McAdams%20Condorcet%20Jury.PDF [23.2.2006]

Dhillon G (1997) *Managing Information System Security*, MacMillan Press ltd, London

Drahos P with Braithwaite J (2003) *Information Feudalism. Who Owns the Knowledge Economy?*, The New Press, New York

Eckhoff T (1983) *Statens Styrningsmuligheter. Særlig i Ressurs- og Miljøspørsmål*, Tanum-Norli, Oslo

Egeskov C and Christensen JA (2003) Behovet for forsikring af software-producentens mulige erstatningansvar, *Nordisk Försäkringstidskrift* (Scandinavian Insurance Quarterly), 2003(1): 51-73

Eijlander P (2005) Possibilities and Constraints in the Use of Self-Regulation and Co-Regulation in Legislative Policy: Experience in the Netherlands – Lessons to Be Learned for the EU? *Electronic Journal of Comparative Law*, Volume 9.1 (January 2005), available at http://www.ejcl.org/91/art91-1.html [8.2.2006]

Eisenberg MA (2003) Mistake in Contract Law, *California Law Review*, 91(6): 1575-1644

Eisenberg MA (2003) Disclosure in Contract Law, *California Law Review*, 91(6): 1647-1692

EFF, Electronic Frontier Foundation (2003) Unintended Consequences: Five Years under the DMCA, v. 3, September 24, 2003, available at http://www.eff.org/IP/DMCA/?f=unintended_consequences.html [14.2.2006]

Ellickson R (2001) The Market for Social Norms, *American Law and Economics Review*, 3(1): 1-49

Ellmer E, Merkl D, Quirchmayr G and Min Tjoa A (1996) Process Model Reuse to Promote Organizational Learning in Software Development, in Proceedings of the 20th Annual Int'l Computer Software and Applications Conference (COMPSAC'96), Seoul, Korea, August 19-20, IEEE Press, p. 21-26, available at http://www.ifs.tuwien.ac.at/ifs/research/pub_ps/ell_compsac96.ps.gz

Ely JC and Välimaki J (2003) Bad reputation, *The Quarterly Journal Of Economics*, 118(3): 785-814

Ervasti K and Tala J (1996) *Lainvalmistelu ja vaikutusten ennakointi*, Edita, Helsinki

Etzioni A (2000) Social Norms: Internalization, Persuasion, and History, *Law & Society Review*, 34(1): 157-178

Evan WM (1980) Law as an Instrument of Social Change, in Evan WM (ed.) *The Sociology of Law. A Social-Structural Perspective*, The Free Press, New York, p. 554-563. Reprinted from Gouldner AW and Miller SM (eds., 1965) *Applied Sociology: Opportunities and Problems*, The Free Press, New York, p. 285-293

Feddersen TJ and Gilligan TW (2001) Saints and Markets: Activists and the Supply of Credence Goods, *Journal of Economics & Management Strategy*, 10(1): 149–171

Fehr E and Falk A (2002) Psychological foundations of incentives, Schumpeter Lecture, Annual Conference of the European Economic Association 2001, *European Economic Review*, 46(4-5): 687-724Feld WJ and Jordan RS (1988) *International Organizations. A Comparative Approach*, Praeger, New York

Ferrejoli L (2001) Fundamental Rights, *International Journal for the Semiotics of Law*, 14(1): 1-33

Fine GA (2001) Enacting Norms: Mushrooming and the Culture of Expectations and Explanations, in Hechter M and Opp K-D (eds.) *Social Norms*, Russel Sage Foundation, New York, p. 139-165

Fischhoff B and Merz JF (1994) The Inconvenient Public: Behavioral Research Approaches to Reducing Product Liability Risks, in Hunziker JR and Jones TO (eds.) *Product Liability and Innovation: Managing Risk in an Uncertain Environment*, National Academy Press, Washington, DC, p. 159-189 http://www.nap.edu/books/0309051304/html/ [23.2.2006]

Fisher D (2002) Contracts Getting Tough on Security, *eWeek*, April 15, 2002, available at http://www.eweek.com/article2/0,1895,1658531,00.asp [29.3.2006]

Froomkin AM (2003) Habermas@Discourse.net: Toward a Critical Theory of Cyberspace, *Harvard Law Review*, 116(3): 749-873, also available at http://osaka.law.miami.edu/~froomkin/discourse/ils.pdf [23.2.2006]

Fuggetta A (2000) Software Process: A Roadmap, in Finkelstein A (ed.) *The Future of Software Engineering, 22nd International Conference on Software Engineering*, ACM, New York, June

Gal-Or E and Ghose A (2003) *The Economic Consequences of Sharing Security Information*, paper presented at the Second Annual Workshop on Economics and Information Security, Robert H. Smith School of Business, University of Maryland, May 29-30, 2003, http://www.cpppe.umd.edu/rhsmith3/papers/Final_session7_galor.ghose.pdf [21.2.2006]

Garland D (1990) *Punishment and Modern Society: A Study in Social Theory*, University of Chicago Press, Chicago

Gavison R (1991) Comment: Legal Theory and the Role of Rules, *Harvard Journal of Law & Public Policy*, 14(3): 727-771

Gehring RA (2002) *Software development, Intellectual Property Rights, and IT Security*, paper presented at the First Workshop on Economics and Information Security, University of California, Berkley, May 16-17, 2002, available at http://www.sims.berkeley.edu/resources/affiliates/workshops/econsecurity/econws/44.pdf [16.2.2006]

Gehring RA (2001) *"Software Patents" — IT–Security at Stake?*, Paper presented at the international conference "Innovations for an e–Society. Challenges for Technology Assessment", 17–19 October, Berlin, Germany, available at http://ig.cs.tu-berlin.de/oldstatic/ap/rg/2000-05/2001-10/Gehring2001Full-SWPatITSec.pdf [21.2.2006]

Geistfeld M (1999) Products Liability, in Bouckaert B and De Geest G (eds., 2000) *Encyclopedia of Law and Economics, Volume III. The Regulation of Contracts*, Edward Elgar, Cheltenham, also available at http://allserv.rug.ac.be/~gdegeest/5140book.pdf [21.2.2006]

Gomulkiewicz RW and Williamson ML (1996) A Brief Defense of Mass Market Software License Agreements, *Rutgers Computer & Technology Law Journal*, 12(2):335-369

Gordon LA, Loeb MP, Lucyshyn W and Richardson R (2005) *2005 CSI/FBI Computer Crime and Security Survey*, Computer Security Institute (CSI) publications, available at http://www.gocsi.com/ [23.2.2006]

Graff MG and van Wyk KR (2003) *Secure Coding. Principles & Practice*, O'Reilly & Associates, Sebastopol, USA

Granick JS (2004) The Price of Restricting Vulnerability Publications, *International Journal of Communications Law and Policy* (IJCLP), Issue 9, part 2, http://www.digital-law.net/IJCLP/Cy_2004/ijclp_webdoc_10_Cy_2004.htm [23.2.2006]

Gunningham N and Grabovsky P (1998) *Smart Regulation: Designing Environemental Policy*, Oxford University Press, Oxford

Habermas J (1996) *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, Polity Press, Cambridge (translated by William Rehg)

Harju H (2002) *Kustannustehokas ohjelmiston luotettavuuden suunnittelu ja arviointi*, Osa 1, [Costeffective design and assessment of dependable software, Part 1], VTT Tiedotteita – Research Notes 2151, Valtion teknillinen tutkimuskeskus, Espoo, http://virtual.vtt.fi/inf/pdf/tiedotteet/2002/T2151.pdf [6.3.2006] (in Finnish)

Harju H and Koskela M (2003) *Kustannustehokas ohjelmiston luotettavuuden suunnittelu ja arviointi*, Osa 2 [Cost-effective reliability design and assessment of software, Part 2], VTT Tiedotteita – Research Notes 2193,Valtion teknillinen tutkimuskeskus, Espoo, http://virtual.vtt.fi/inf/pdf/tiedotteet/2003/T2193.pdf [6.3.2006](in Finnish)

Harmathy A (1987) The Influence of Legal Systems on Modes of Implementation of Economic Policy, in Daintith T (ed.) *Law as an Instrument of Economic Policy: Comparative and Critical Approaches*, European University, Series A, No. 7, Walter de Gruyter, Berlin, p. 245-267

Hart HLA (1994) *The Concept of Law*, 2nd ed. (with a new Postscript), Clarendon Press, Oxford

Harrold MJ (2000) Testing: A Roadmap, in Finkelstein A (ed.) *The Future of Software Engineering, 22nd International Conference on Software Engineering*, ACM, New York, June, p. 61-73

Havana T (2003) *Communication in the Software Vulnerability Reporting Process*, M.A. thesis, University of Jyväskylä, http://www.ee.oulu.fi/research/ouspg/protos/sota/reporting/gradu.pdf [23.2.2006]

Havana T and Röning J (2003) Communication in the Software Vulnerability Process, *Proceedings of the 15th FIRST Conference on Computer Security Incident Handling*, Ottawa, Canada, June 22-27, 2003, http://www.ee.oulu.fi/research/ouspg/protos/sota/FIRST2003-communication/paper.pdf [23.2.2006]

Hechter M and Opp K-D (2001) What Have We Learned about the Emergence of Social Norms, in in Hechter M and Opp K-D (eds.) *Social Norms*, Russel Sage Foundation, New York, p. 394-417

Helenius M (2005) *Tietoturvallisuuden tutkimus ja opetus. Nykytilanne ja kehittämismahdollisuudet*, Tietoyhteiskuntainstituutin raportteja 2/2005, Tampereen yliopisto, Tampere

Hellner J (1990) *Lagstiftning inom förmögenhetsrätten. Praktik, teori och teknik*, Juristförlaget, Stockholm

Hellner J (1985) Legislation and Sociology: The Law of Torts, in Kivivuori A (ed.) *Law Drafting and Sociology*, Ministry of Justice, Law Drafting Department, Publication Series No. 2/1985, Helsinki, p. 45-67

Hemmo M (1992) Kuluttajamainonnan informatiivisuusvaatimuksista, *Lakimies*, the Journal of the Finnish Lawyers' Association, 90(3): 368-371

Hemmo M (2005) *Oikeudellisen riskienhallinnan perusteita*, Helsingin yliopiston oikeustieteellisen tiedekunnan julkaisut, Forum Iuris, Helsinki

Hemmo M (2004a) Sopimuksiin liittyvät vastuuriskit, in Aalto-Setälä I, Amper M, Haussila P, Hemmo M, Lintumaa S, Saloheimo J, Salomaa P, Soikkeli L, Strömberg H, Tuomainen J, and Virtanen P (2004) *Yrityksen ja yhteisön vastuuriskit. Oikeudellisen riskienhallinnan perusteet*, 2nd ed., Tietosanoma Oy, Helsinki, p. 13-31

Hemmo M (1998) *Sopimus ja delikti. Tutkimus vahingonkorvausoikeuden vastuumuodoista*, Kauppakaari, Lakimiesliiton Kustannus, Helsinki

Hemmo M (2003a) *Sopimusoikeus I*, 2nd ed., Talentum, Helsinki

Hemmo M (2003b) *Sopimusoikeus II*, 2nd ed., Talentum, Helsinki

Hemmo M (2005) *Sopimusoikeus III*, Talentum, Helsinki

Hemmo M (2004b) Tuotevastuuriskit, in Aalto-Setälä I, Amper M, Haussila P, Hemmo M, Lintumaa S, Saloheimo J, Salomaa P, Soikkeli L, Strömberg H, Tuomainen J, and Virtanen P (2004) *Yrityksen ja yhteisön vastuuriskit. Oikeudellisen riskienhallinnan perusteet*, 2nd ed., Tietosanoma Oy, Helsinki, p. 31-49

Hemmo M (2002) *Vahingonkorvausoikeuden oppikirja*, WSOY Lakitieto, Helsinki

Hemmo M (1996) *Vahingonkorvauksen sovittelu ja moderni korvausoikeus*, [The Reduction of Damages and Modern Compensation Law] Suomalaisen lakimiesyhdistyksen julkaisuja, A-sarja, No. 209, Suomalainen lakimies-yhdistys, Helsinki, with an English summary

Hemmo M (1999) Vuoden 2000 ongelma ja siviilioikeus [The year 2000 problem and civil law], *Oikeustiede – Jurisprudentia XXXII*, Suomalaisen lakimiesyhdistyksen vuosikirja, Gummerus Kirjapaino Oy, Jyväskylä, p. 5-80, with an English summary

den Hertog JA (2000) General Theories of Regulation, in Bouckaert B and De Geest G (eds., 2000) *Encyclopedia of Law and Economics, Volume III. The Regulation of Contracts*, Edward Elgar, Cheltenham, p. 223-270, http://allserv.rug.ac.be/~gdegeest/5000book.pdf [22.2.2006]

van Hoecke M (2002) *Law as Communication*, Hart Publishing, Oxford and Portland, Oregon

Hood C (1986) *The Tools of Government*, Chatham House, Chatham, New Jersey

Hood C and Scott C (2000) Regulating Government in a 'Managerial' Age: towards a cross-national perspective, *CARR Discussion Paper*, no. 1, October, http://www.lse.ac.uk/collections/CARR/pdf/Disspaper1.pdf [16.2.2006]

Hood C, Rothstein H and Baldwin R (2001) *The Government of Risk. Understanding Risk Regulation Regimes*, Oxford University Press, Oxford

Horne C (2001) Sociological Perspectives on the Emergence of Norms, in Hechter M and Opp K-D (eds.) *Social Norms*, Russel Sage Foundation, New York, p. 3-35

Hosein IR (2003) *Regulating the Technological Actor: How Governments Tried to Transform the Technology and the Market for Cryptography and the Implications for the Regulation of Information and Communications Technologies*, London School of Economics, Department of Information Systems, submitted for a PhD in Information Systems, http://www.lse.ac.uk/collections/informationSystems/pdf/theses/hosein2.pdf [23.2.2006]

Hovav A and D'Arcy J (2005) Capital market reaction to defective IT products: The case of computer viruses, *Computers & Security* 24(5): 409-424

Howells G (2001) The Millennium But and Product Liability, in Wilhelmsson T, Tuominen S and Tuomola H (eds.) *Consumer Law in the Information Society*, Kluwer Law International, The Hague, p. 295-307. Originally published in the Journal of Information, Law and Technology (JILT) 1999(2), available at http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/1999_2/howells/ [27.3.2006]

Howells G and Wilhelmsson T (1997) EC and US Approaches to Consumer Protection – Should the Gap be Bridged?, *Yearbook of European* Law, 17: 207-268

Hugenholtz B (2000) Why the Copyright Directive is Unimportant, and Possibly Invalid, *European Intellectual Property Review* (EIPR) 11: 501-502, available also at http://www.ivir.nl/publications/hugenholtz/opinion-EIPR.html [published 11.10.2000, accessed 14.2.2006]

Hunt A (1997) The Politics of Law and the Law of Politics, in Tuori K, Bankowski Z and Uusitalo J (eds.) *Law and Power. Critical and Socio-Legal Essays*, Deborah Charles Publications, Liverpool, p. 51-83

Hunziker JR and Jones TO (eds., 1994) *Product Liability and Innovation: Managing Risk in an Uncertain Environment*, Washington DC, National Academy Press, http://www.nap.edu/books/0309051304/html/ [14.2.2006]

Hydén H (2001) *Rättsregler. En introduktion till juridiken*, Studentlitteratur, Lund

Häyhä J (1999) Ankara vastuu ja vahingonkorvausoikeuden järjestelmä, Oikeustiede – Jurisprudentia XXXII, Suomalaisen lakimiesyhdistyksen vuosikirja, Gummerus Kirjapaino Oy, Jyväskylä, p. 81-150, with an English summary

Irlenbusch B (2004) Relying on a man's word? An experimental study on non-binding contracts, *International Review of Law and Economics*, 24(3): 299–332

Jarass HD (1988) Regulation as an Instrument of Economic Policy, in Daintith T (ed.) *Law as an Instrument of Economic Policy: Comparative and Critical Approach*, Walter de Gruyter, Berlin, p. 75-97

Joerges C, Falke J, Miclitz HW and Brüggemeier G (1991) *European Product Safety, Internal Market Policy and the New Approach to Technical Harmonisation and Standards*, EUI Working Paper LAW, Nos. 91/10-14, European University Institute, Florence, available at http://www.iue.it/LAW/WP-Texts/Joerges91/ [13.2.2006]. Original German version: Die Sicherheit von Kosnumgütern und die Entwicklung der Gemeinschaft, Nomos, 1988

Joerges C, Schepel H and Vos E (1999) The Law's Problems with the Involvement on Non-Governmental Actors in Europe's Legislative Process: The Case of Standardisation under the 'New Approach', EUI Working Paper Law, No. 99/9, European University Institute, San Domenico, available at http://cadmus.iue.it/dspace/retrieve/948/law99_9.pdf [14.2.2006]

Jolls C, Sunstein CR and Thaler RH (1998) A Behavioral Approach to Law and Economics, *Stanford Law Review* 50(5): 1471-1551, available also at http://www.law.harvard.edu/programs/olin_center/papers/pdf/236.pdf [14.2.2006]

Kaisanlahti T (1999) Riskin pulverointi vahingonkorvausoikeuden tehtävänä, in Kanniainen V and Määttä K (eds.) *Näkökulmia oikeustaloustieteeseen 3*, Kauppakaari OYJ, Lakimiesliiton Kustannus, Helsinki 1999, p. 85-105

Kaner C (1997) The Impossibility of Complete Testing, at http://www.kaner.com/pdfs/imposs.pdf [23.2.2006] published in Software QA, 4(4)

Kaner C (2000) *Software Engineering and UCITA*, available at http://www.badsoftware.com/engr2000.htm, last modified September 16, 2000 [2.1.2006] initially published in *John Marshall Journal of Computer and Information Law*, 18(2): 435-546, Winter 2000.

Kaner C and Pels D (1997) Software Customer Dissatisfaction, at http://www.kaner.com/pdfs/sqastat2.pdf [23.2.2006] published at *Software QA*, 4(3)

Kannan K and Telang R (2004) An Economic Analysis of Market for Software Vulnerabilities, paper presented at *The Third Annual Workshop on Economics and Information Security* (WEIS04), May 13-14, 2004, University of Minnesota, Minneapolis, USA http://www.dtc.umn.edu/weis2004/kannan-telang.pdf [23.2.2006]

Karlsson J and Ryan K (1997) A Cost-Value Approach for Prioritizing Requirements, *IEEE Software*, 14(5): 67-75

Kaspersen HW (1994) Foreword, in Sizer R, Yngström L, Kaspersen H, Fischer-Hübner S (eds.) *Security and Control of Information Technology in Society*, Proceedings of the IFIP TC9/WG9.6 Working Conference on Security and Control of Information Technology in Society on Board M/S Illich and ashore at St. Petersburg, Russia, 12-17 August, 1993, IFIP Transaction, A-43, North-Holland, p. 7-10

Kaspersen HW (1992) How to Advance Computer Security by Legal Instruments?, in Kilian W and Wiebe A (eds.) *Data Security in Computer Networks and Legal Problems*, S. Toeche-Mittler Verlag, Darmstadt, Proceedings of a Working Conference in Hannover/Germany on September 23-24, 1991, p. 85-95

Klami HT (1977) *Oikeudellisen sääntelyn yleinen teoria*, Turun yliopiston yksityisoikeuden laitoksen julkaisuja 12/1977, Turku

Keil M and Carmel E (1995) Customer-developer links in software development, *Communications of the ACM*, 38(5): 33–44

Kesan JP, Majuca RP and Yurcik WJ (2005) The Economic Case for Cyberinsurance, *University of Illinois Law & Economics Research Paper* No. LE04-004, paper presented at a Stanford Law School Symposium: Securing Privacy in the Internet Age, March 13-14 2004, Stanford Law School, CA, USA, http://ssrn.com/abstract=577862 [23.2.2006]

Keskitalo P (2000) *From Assumptions to Risk Management. An Analysis of Risk Management for Changing Circumstances in Commercial Contracts, Especially in the Nordic Countries*, Kauppakaari Oyj, Lakimiesliiton Kustannus, Helsinki

Kilian W (1992) Data Security in Computer Networks and Legal Problems, in Kilian W and Wiebe A (eds.) *Data Security in Computer Networks and Legal Problems*, S. Toeche-Mittler Verlag, Darmstadt, Proceedings of a Working Conference in Hannover/Germany on September 23-24, 1991

Kivivuori A (2005) Vahingonkorvausvastuun tarkoitusperät, in Halila H, Hemmo M and Sisula-Tulokas L (eds. 2005) *Juhlajulkaisu Esko Hoppu 1935 –15/1 – 2005*, Suomalainen Lakimiesyhdistys, Helsinki, p. 163-173

Kunreuther H. and Heal G (2003) Interdependent Security, *Journal of Risk and Uncertainty*, 26(2-3): 231-249, also available at http://opim.wharton.upenn.edu/risk/downloads/02-06-HK.pdf [21.2.2006]

Kunreuther H, Heal G and Orszag PR (2002) Interdependent Security: Implications for Homeland Security Policy and Other Areas, *The Brookings Institution Policy Brief*, number108, October, http://www.brookings.org/dybdocroot/comm/policybriefs/pb108.pdf [23.2.2006]

Kuvaja P, Maansaari J, Seppänen V and Taramaa J (1999) Specific Requirements for Assessing Embedded Product Development, in Oivo M and Kuvaja P (eds.) *Proceedings of the International Conference on Product Focused Software Process Improvement*, Oulu, Finland, June 22-24, 1999, Technical Research Center of Finland, Espoo

Laakso M, Takanen A and Röning J (2001) Introducing Constructive Vulnerability Disclosures, *Proceedings of the 13th FIRST Conference on Computer Security Incident Handling*, Toulouse, June 17-22, 2001, http://www.ee.oulu.fi/research/ouspg/protos/sota/FIRST2001-disclosures/ [updated 26.7.2001, visited 23.2.2006]

Landwehr C (2002) *Improving Information Flow in the Information Security Market*, paper presented at the Workshop on Economics and Information Security, University of California, Berkley, May 16-17, 2002, http://www.sims.berkeley.edu/resources/affiliates/workshops/econsecurity/econws/11.doc [21.2.2006]

Latour B (2000) When Things Strike Back: A Possible Contribution of Science Studies to the Social Sciences, *British Journal of Sociology*, 51(1): 107-124

Lemley MA and McGowan D (1998) *Legal Implications of Network Economic Effects*, http://papers.ssrn.com/paper.taf?ABSTRACT_ID=32212 [21.2.2006] published in Cal. L. Rev. 86: 479

Lessig L (1999) *Code and Other Laws of Cyberspace*, Basic Books, New York

Lessig L (1999) The Law of the Horse. What Cyberlaw Might Teach, *Harvard Law Review*, 113(2): 501-550

Lessig L (1998) The New Chicago School, *Journal of Legal Studies*, 27(2): 661-691, also available at http://lessig.org/content/articles/works/LessigNewchicschool.pdf [23.2.2006]

Lessig L (1995) The Regulation of Social Meaning, *University of Chicago Law Review*, 62(3): 944-1045, also available at http://www.lessig.org/content/articles/works/regulation-socialmeaning.pdf [23.2.2006]

Levin DB (2002) Student Note: Building social norms on the Internet, *Yale Journal of Law & Technology*, 4(2001-2002): 97-138

Lewis JA (2005) Aux Armes, Citoyens: Cyber Security and Regulation in the United States, *Telecommunications Policy*, 29(11): 821-830

Liebowitz SJ and Margolis SE (1994) Network Externality: An Uncommon Tragedy, *Journal of Economic Perspectives*, 8(2), also available at http://wwwpub.utdallas.edu/~liebowit/jep.html [21.2.2006]

Lindberg A and Westman D (1999) *Praktisk IT-rätt*, 2nd ed., Norstedts Juridik AB, Stockholm

Liska AE (1997) Modelling the Relationships Between Macro Forms of Social Control, *Annual Review of Sociology*, 23(1): 39-61

Lloyd IJ (2005) *Information Technology Law*, 4th ed., Oxford University Press, Oxford

Lookabaugh T and Sicker DC (2003) *Security and Lock-In: The Case of the U.S. Cable Industry*, paper presented at Economics and Information Security Workshop, University of Maryland, May 29-30, http://www.cpppe.umd.edu/rhsmith3/papers/Final_session8_lookabaugh.sicker.pdf [21.2.2006]

Lundvall BÅ (2000) Understanding the Role of Education in the Learning Economy. The Contribution of Economics, in OECD-CERI, *Knowledge Management in the Learning Society*, pp. 11-35, Paris, OECD, pp. 18-21

MacCormack A (2001) Product Development Practices that Work: How Internet Companies Build Software, *Sloan Management Review*, 42(2): 75-84

MacCormack A, Verganti R and Iansiti M (2001) Developing Products on "Internet Time": The Anatomy of a Flexible Development Process, *Management Science*, 47(1): 133-150.

McCormick N (1981) *H.L.A. Hart*, Jurists: Profiles in Legal Theory, Stanford University Press, Stanford

MacCormick N (1983) On Legal Decisions and their Consequences: from Dewey to Dworkin, *New York University Law Review*, 58(2): 239-258, republished in Aarnio A and MacCormick N (eds., 1992) *Legal Reasoning. Volume II*, The International Library of Essays in Law and Legal Theory, Dartmouth, Aldershot, p. 83-102

MacDonald (1999) Y2K and Contractual Exemption Clauses, *Journal of Information, Law and Technology* (JILT), 1999(2), available at http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/1999_2/macdonald/ [27.3.2006]

Mackaay E (1982) *Economics of Information and Law*, Kluwer, Nijhoff Publishing, Dordrecht

Mackaay E (1992) The Public's Right to Information, in Korthals Altes WF, Dommering EJ, Hugenholtz PB, and Kabel JJC (eds.) *Information Law Towards the 21st Century*, Kluwer, Deventer

Mackaay E and Leblanc V (2003) *The Law and Economics of Good Faith in the Civil Law of Contract*, paper prepared for the 2003 Conference of the European Association of Law and Economics, at Nancy, France, 18-20 September 2003, http://www.cdaci.umontreal.ca/pdf/mackaay_law_economics.pdf [23.2.2006]

MacKenzie D and Wajcman J (1987) Introductory Essay: The Social Shaping of Technology, in MacKenzie D and Wajcman J (ed.) *The Social Shaping of Technology. How the Refrigerator got its Hum*, Milton Keynes Open University Press, Philadelphia, p. 3-28

Mathiesen T (2005) *Rätten i samhället. En introduktion till rättssociologin*, femte upplagan, Studentlitteratur, Lund

Matthews S and Postlewaite A (1985) Quality Testing and Disclosure, *Rand Journal of Economics*, 16(3): 328-340

Mayntz R (1983) The Conditions of Effective Public Policy – A New Challenge for Policy Analysis, *Policy and Politics*, 11(2): 123-145

Mayntz R (1988) Political Intentions and Legal Measures: The Determinants of Policy Decisions, in Daintith T (ed.) *Law as an Instrument of Economic Policy: Comparative and Critical Approach*, Walter de Gruyter, Berlin

Mazmanian DA and Sabatier PA (1989) *Implementation and Public Policy*, University Press of America, Lanham

Mattei U (1998) *Comparative Law and Economics*, The University of Michigan Press, Ann Arbor

McAdams RH (1997) The Origin, Development, and Regulation of Norms, *Michigan Law Review*, 96(2): 338-434

McAdams RH and Nadler J (2004) A Third Model of Legal Compliance: Testing for Expressive Effects in a Hawk/Dove Game, *International Association for Conflict Management (IACM) 17th Annual Conference Paper No. P-107*, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=573582 [23.2.2006]

McNutt P (2000) Public Goods and Club Goods, in Bouckaert B and De Geest G (eds.) *Encyclopedia of Law and Economics, Volume I. The History and Methodology of Law and Economics*, Edward Elgar, Cheltenham, p. 927-951, http://allserv.rug.ac.be/~gdegeest/0750book.pdf [21.2.2006]

Meijboom AP (1989) Legal Rights to Source Code, in Vandenberghe GPV (ed.) *Advanced Topics of Law and Information Technology*, Kluwer Law and Taxation Publishers, Boston, p. 105-127

Meltzer J, Freeman R and Thomson S (2003) *Product Liability in the European Union. A Report for the European Commission*, MARKT/2001/11/D, available at the web pages of the European Commission at http://europa.eu.int/comm/enterprise/regulation/goods/liability_en.htm [22.2.2006]

Mercuro N and Medema SG (1997) *Economics and the Law. From Posner to Post-Modernism*, Princeton University Press, Princeton, New Jersey

Miettinen T (2001) *Tieteen vapaus: Julkisoikeudellinen tutkimus tieteenharjoittajan itsemääräämisoikeudesta, tieteen itsekontrollista ja yliopiston itsehallinnosta*, Kauppakaari, Helsinki

Migdal A (1999) Shrinkwrap Licenses Abroad, *Journal of Internet Law*, Vol. 2, partly reprinted in Lemley MA, Menell PS, Merges RP, and Samuelson P (2000) *Software and Internet Law*, Aspen Law & Business, New York, p. 751-753

Mononen M (2004) *Yritysten välinen tuotevastuu*, Talentum Media Oy, Helsinki

Moore MS (1989) Authority, Law, and Razian Reasons, *Southern California Law Review*, 62: 827-897

Moore MS (1991) Three concepts of rules, *Harvard Journal of Law & Public Policy*, 14(3): 771-797

Moran M and Prosser T (eds., 1994) Privatization and Regulatory Change in Europe, Open University Press, Buckingham

Morawetz T (1996) Law and Literature, in Patterson D (ed.) *A Companion to Philosophy of Law and Legal Theory*, Blackwell Publishers, Oxford, p. 450-461

Mundie C (2002) *Security: Source Access and the Software Ecosystem*, paper presented in Open Source Software: Economics, Law and Policy, Roundtable, June 20, Toulouse, France, June 20-21, available at http://idei.fr/doc/conf/sic/papers_2002/mundie.pdf [21.2.2006]

Natt och Dag J (2002) *Elicitation and Management of User Requirements in Market-Driven Software Development*, LUCAS Technical Report 146, Lund University: Lund Institute of Technology, Lund, http://www.lucas.lth.se/publications/pub2002/020612Johan.pdf [16.2.2006]

Niemivuo M (2002) *Kansallinen lainvalmistelu* [Finnish National Legislative Drafting], Kauppakaari: Lakimiesliiton Kustannus, Helsinki, (with an English Summary)

Nissenbaum H (2001) How Computer Systems Embody Values, *IEEE Computer*, 34(3): 118-120

Noll J (2004) Comparing Quality Signals as Tools of Consumer Protection: Are Warranties Always Better than Advertisements to Promote Higher Product Quality?, *International Review of Law and Economics*, 24(2): 227–239

Noll J (2003) Does One Size Fit All? A Note on the Harmonization of National Warranty Law as a Tool of Consumer Protection, *European Journal of Law and Economics*, 16(2): 219-231

Odlyzko AM (1998) Smart and stupid networks: Why the Internet is like Microsoft, *ACM netWorker*, 2(5): 38-46, also available at http://www.acm.org/networker/issue/9805/ssnet.html [21.2.2006]

Ogus AI (1994) *Regulation: Legal Form and Economic Theory*, Oxford University Press, Oxford

Opp K-D (1979) The Emergence and Effects of Social Norms. A Confrontation of Some Hypotheses of Sociology and Economics, *Kyklos, International Review for Social Sciences*, 32(4): 775-802

O'Regan G (2002) *A Practical Approach to Software Quality*, Springer, New York

OIS, Organization for Internet Safety (2004) *Guidelines for Security Vulnerability Reporting and Response*, Version 2.0, 1st September 2004, available at http://www.oisafety.com/guidelines/ [31.10.2005]

OWASP, Open Web Application Security Project, Secure Software Development Contract Annex, available at http://www.owasp.org/documentation/legal.html [29.3.2006]

Parker C (1999) *Just Lawyers*, Oxford University Press, Oxford

Parker C, Scott C, Lacey N and Braithwaite J (2004) Introduction, in Parker C, Scott C, Lacey N and Braithwaite J (eds.) *Regulating Law*, Oxford University Press, Oxford

Parsons T, Shils EA, Allport GW, Kluckhohn C, Murray HA, Sears RR, Sheldon RC, Stouffer SA and Tolman EC (1951) Some Fundamental Categories of the Theory of Action: A General Statement, in Parsons T and Shils EA (eds.) *Toward a General Theory of Action: Theoretical Foundations for the Social Sciences*, New York and Evanston

Personick SD and Patterson CA (eds. 2003) *Critical Information Infrastructure Protection and the Law: An Overview of Key Issues*, Committee on Critical Information Infrastructure Protection and the Law, Computer Science and Telecommunications Board, National Academy of Engineering, National Research Council of the National Academies, The National Academies Press, Washington, D.C., http://bob.nap.edu/html/ciip/ [2.2.2006]

Peters BG (2002) The Politics of Tool Choice, in Salamon LM (ed.) *The Tools of Government. Guide to the New Governance*, Oxford University Press, New York, p. 552-565

Pipkin DL (2000) *Information Security. Protecting the Global Enterprise*, Prentice Hall PTR, New Jersey

Polinsky AM and Rogerson WP (1983) Products liability, consumer misperceptions, and market power, *Bell Journal of Economics*, 14(2): 581-589

Post DG (2000) What Larry Doesn't Get: A Libertarian Response to Lessig's Code and Other Laws of Cyberspace, *Stanford Law Review*, 52: 1439-1459, also available at http://www.temple.edu/lawschool/dpost/Code.pdf [23.2.2006]

Postema GJ (1991) Positivism, I presume?...comments on Schauer's `Rules and the Rule of Law', *Harvard Journal of Law & Public Policy*, 14(3): 797-823

Potts C (1995) Invented Requirements and Imagined Customers: Requirements Engineering for Off-the-Shelf Software, *Proceedings of Second IEEE International Symposium on Requirements Engineering* (RE'95), IEEE Computer Society Press, York, UK, also available at http://www.cc.gatech.edu/fac/Colin.Potts/pubs/1995/re95/re95InventedReqtsOTS.pdf [23.2.2006]

Pöyhönen J (2002) The Law of Obligations, in Pöyhönen J (ed.) *An Introduction to Finnish Law*, Kauppakaari, Finnish Lawyers's Publishing, Helsinki, p. 59-111

Pöyhönen J (2001) Törkeä tuottamus ja vastuunrajoitusehdot, in Arvanne-Potrykus H et al. (eds.) *Asianajotoimisto Borenius & Kemppinen 90 vuotta*, Gummerus, Jyväskylä, p. 73-97

Pöyhönen J (2000) *Uusi varallisuusoikeus*, Kauppakaari Oyj, Lakimiesliiton Kustannus, Helsinki

Pöysti T (2001) *ENLIST Information Security Commentary*, available at http://www.ulapland.fi/home/oiffi/enlist/commentary/information_security.html [23.2.2006]

Pöysti T (2004) ICT and Legal Principles: Sources and Paradigm of Information Law, in Wahlgren P (ed.) *IT Law*, Scandinavian studies in law 47, Stockholm Institute for Scandinavian Law, Stockholm, p. 560-600

Pöysti T (1999) *Tehokkuus, informaatio ja Eurooppalainen oikeusalue*, Hakapaino Oy, Helsinki

Quirchmayr G, Slay J, Koronios A and Darzano K (2002) A Business Process Engineering Based Approach towards Incorporating Security in the Design of Global Information Systems, in *Proceedings of the 7th Pacific Asia Conference on Information Systems (PACIS 2003)*, 10-13 July 2003, Adelaide, Sourth Australia, paper nro. 77, p. 1114-1120, available at http://www.pacis-net.org/file/2003/papers/it-is-infrastructure/280.pdf [3.3.2006]

Rajala R, Rossi M, Tuunainen VK and Korri S (2001) Software Business Models. A Framework for Analyzing Software Industry, *TEKES Technology Review 108/2001*, Paino-Center Oy, Helsinki

Raz J (1979) *The Authority of Law: Essays of Law and Morality*, Oxford University Press, Oxford

Raz J (1973) On the Functions of Law, in Simpson AW (ed.) *Oxford Essays in Jurisprudence (Second Series)*, Clarendon Press, Oxford, p. 278-305

Raz J (1975) Reasons for Action, Decisions and Norms, *Mind*, New Series, 84(336): 481-499

Redondo C (1999) *Reasons for Action and the Law*, Kluwer Academic Publishers, Dordrecht

Reich N (1985) Social Sciences and Law Reform: Experiences of Regulating Contract Law in the Consumer Interest, in Kivivuori A (ed.) *Law Drafting and Sociology*, Ministry of Justice, Law Drafting Department, Publication Series No. 2/1985, Helsinki, p. 67-107

Reidenberg JR (1998) Lex Informatica. The Formulation of Information Policy Rules through Technology, *Texas Law Review*, 76(3): 553-584

Reifer DJ, Boehm BW, Gangadharan M (2003) Estimating the Cost of Security for COTS Software, in Erdogmus H and Weng T (Eds.) *Proceedings of the Second International Conference on COTS-Based Software Systems*, ICCBSS 2003, Ottawa, Canada, February 10-12, 2003, p. 178-186. Published in *Lecture Notes in Computer Science*, Volume 2580/2003, Springer-Verlag, Berlin Heidelberg

Reimann M (2003) Product Liability in a Global Context: the Hollow Victory of the European Model, *European Review of Private Law*, 11(2): 128-155

Riesenhuber K (2001) Party Autonomy and Information in the Sales Directive, in Grundmann S, Kerber W, and Weatherill S (eds.) *Party Autonomy and the Role of Information in the Internal Market*, Walter de Gruyter, Berlin

Rocher G (1998) L'effectivité du droit, in Lajoie A, MacDonald RA, Janda R and Rocher G (1998) *Théories et émergence du droit : pluralisme, surdétermination et effectivité*, Éditions Thémis, Montreal

Rothchild J (2001) Co-Regulating the Internet, in Wilhelmsson T, Tuominen S and Tuomola H (eds.) *Consumer Law in the Information Society*, Kluwer Law International, The Hague, p. 179-205

Rubin EL (2000) Legal Scholarship, in Patterson D (ed.) *A Companion to Philosophy of Law and Legal Theory*, Blackwell Companions to Philosophy, Blackwell Publishers, Oxford, p. 562-573

Rubin PH (2000) Information Regulation (incl. Regulation of Advertising), in Bouckaert B and De Geest Gerrit (eds.) *Encyclopedia of Law and Economics, Volume III. The Regulation of Contracts*, Edward Elgar, Cheltenham, p. 271-295,
http://allserv.rug.ac.be/~gdegeest/5110book.pdf [23.2.2006]

Rustad ML and Koenig TH (2005) Harmonizing Cybertort Law for Europe and America, *Journal of High Technology Law*, 5(13): 13-59, available at http://www.jhtl.org/V5N1/04_JHTL_Lambert_RustadKoenig.pdf [15.2.2006]

Råman J (2004) Network Effects and Software Development - Implications for Security, in Sprague RH (ed.) *Proceedings of the 37th Hawaii International Conference on System Sciences* (HICSS-37), 5-8 January 2004, Big Island, HI, USA, Abstracts and CD-ROM of Full Papers, IEEE Computer Society, Los Alamitos, CA

Saarenpää A (1984) Court Decisions as the Focus of Study, in *Scandinavian Studies in Law 1984*, Uppsala, 1984, p. 125-150.

Saarenpää A (2002) Data Security: A Fundamental Right in the e-Society? in *Proceedings of the First International Conference on Electronic Government*, EGOV 2002, Aix-en-Provence, France, September 2-5, 2002, Lecture Notes in Computer Science, Vol. 2456, Springer-Verlag, London, p. 424-429

Saarenpää A (1984) Sääntö, peukalosääntö ja rutiini. Näkökohtia käytännön ja teorian suhteesta, *Oikeustiede – Jurisprudentia* XVII 1984, p. 267-330

Saarenpää A (2005) Tietojenkäsittelystä läsnä-älyyn – katkelmia oikeusinformatiikan kehityksestä, in Tolonen Juha, Annola Vesa and Herler Brita (eds.) *Talousoikeuden taitekohtia, Juhlajulkaisu professori Asko Lehtoselle*, Vaasan yliopiston julkaisuja, Tutkimuksia 266, Vaasa, p. 91-123

Saarenpää A (2002) Yksityisyys, yksityiselämä, yksilön suoja – yksityisyyden käsitteellistä kuvausta, in Haavisto R (ed., 2002) *Professori Kyösti Holman juhlakirja 11.6.2002*, Lapin yliopisto, Rovaniemi, p. 313-337

Saarenpää A and Pöysti T (eds. 1997) *Tietoturvallisuus ja laki. Näkökohtia tietoturvallisuuden oikeudellisesta sääntelystä,* Edita, Helsinki (in Finnish). Executive Summary of the Research Report *Data Security and Law. Perspectives on the Legal Regulation of Data Security*, Group of experts on the regulation of data security, The Finnish Ministry of Finance, The University of Lapland is available in English at http://www.ulapland.fi/home/oiffi/julkaisut/datasec.htm [23.2.2006]

Saarenpää A, Korhonen R and Råman J (2004) *Sähköinen viestintä, tietoturvallisuus ja perusoikeudet*, Report for the National Information Security Advisory Board, Ministry of Transport and Communications Finland, available at the web pages of the Ministry (http://www.mintc.fi) from Viestintätietoa – Tietoturva ja tietosuoja – Kansallinen tietoturvastrategia (in Finnish)

Sabatier PA (1999) The Need for Better Theories, in Sabatier PA (ed.) *Theories of the Policy Process*, Westview Press, Oxford, p. 3-19

Salamon LM (2002) The New Governance and the Tools of Public Action: An Indtroduction, in Salamon LM (ed.) *The Tools of Government. Guide to the New Governance*, Oxford University Press, New York, p. 1-48

Salamon LM (2002) The Tools Approach and the New Governance: Conclusion and Implications, in Salamon LM (ed.) *The Tools of Government. Guide to the New Governance*, Oxford University Press, New York, p. 600-611

Samuelson P and Scotchmer S (2002) *The Law & Economics of Reverse Engineering*, The Yale Law Journal, 111(7):1575-1664

Sand I-J (2000) A Future or a Demise for the Theory of the Sociology of Law: Law as normative, social and communicative function of society, *Retfærd*, 23(3): 55-73

Sandgren C (1996) Om empiri och rättsvetenskap, *Juridisk Tidskrift*, 1995-96(3): 726-749

Sarat A (1980) Support for the Legal System, in Evan WM (ed.) *The Sociology of Law. A Social-Structural Perspective*, The Free Press, New York, p. 167-185. Excerpt from Sarat A (1975) Support for the Legal System: An Analysis of Knowledge, Attitudes, and Behavior, *American Politics Quarterly*, 3: 3-24

Sawyer S (2001) A market-based perspective on information systems development, *Communications of the ACM*, 44(11): 97-103, also available at http://www.acm.org/cacm/1101/p97-sawyer.pdf [16.2.2006]

Sawyer P, Sommerville I and Kotonya G (1999) Improving Market-Driven RE Processes, in Oivo, M and Kuvaja, P (eds.) *Proceedings of the International Conference on Product Focused Software Process Improvement*, Oulu, Finland, June 22-24, Technical Research Center of Finland, Espoo, p. 222-236

Schauer F (1991) *Playing by the Rules. A Philosophical Examination of Rule-Based Decision-Making in Law and in Life*, Clarendon Press, Oxford

Schauer F (1991) The Rules of Jurisprudence: A Reply, *Harvard Journal of Law & Public Policy*, 14(3): 839-853

Schauer F (1991) Rules and the Rule of Law, *Harvard Journal of Law & Public Policy*, 14(3): 645-695

Schneider A and Ingram H (1990) Behavioral Assumptions of Policy Tools, *Journal of Politics*, 52(2): 510-529

Schneider FB (2000) Open source in security: visiting the bizarre, *Proceedings of 2000 IEEE Symposium on Security and Privacy*, IEEE Computer Society, Washington, D.C., USA, p. 126-127

Schneier B (2000) Computer Security: Will We Ever Learn?, *Crypto-Gram Newsletter*, May 15, http://www.counterpane.com/crypto-gram-0005.html [23.2.2006]

Schneier B (2002) Foreword, in Viega J and McGraw G (2002) *Building Secure Software. How to Avoid Security Problems the Right Way*, Addison-Wesley, Boston, USA

Schneier B (2000b) *Secrets & Lies: Digital Security in a Networked World*, John Wiley & Sons Inc., New York

Scholz JT (1991) Cooperative Regulatory Enforcement and the Politics of Administrative Effectiveness, *American Political Science Review*, 85(1): 115-136

Schuck PH (2002) Tort Liability, in Salamon LM (ed.) *The Tools of Government. Guide to the New Governance*, Oxford University Press, New York, p. 466-490

Schwartz GT (1997) Mixed Theories of Tort Law: Affirming Both Deterrence and Corrective Justice, *Texas Law Review*, 75(7): 1801-1834

Scott C (2003) Regulation in the Age of Governance: The Rise of the Post-Regulatory State, *National Europe Centre Paper No. 100*, Australian National University, Canberra, http://www.anu.edu.au/NEC/scott1.pdf [23.2.2006]

Seipel P (1977) *Computing Law: Perspectives on a New Legal Discipline*, LiberFörlag, Stockholm

Senden L (2005) Soft Law, Self-Regulation and Co-Regulation in European Law: Where Do They Meet? *Electronic Journal of Comparative Law*, Volume 9.1 (January 2005), available at http://www.ejcl.org/91/art91-3.html [8.2.2006]

Shah RC and Kesan JP (2003) Incorporating Societal Concerns into Communication Technologies, *IEEE Technology and Society Magazine*, 22(2): 28-33, also available through Social Science Research Network (SSRN) at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=577561 [23.2.2006]

Shapiro C and Varian HR (1999) *Information Rules. A Strategic Guide to the Network Economy*, Harvard Business School Press, Boston, Massachusetts

Shavell S (1992) Liability and the Incentive to Obtain Information about Risk, *The Journal of Legal Studies*, XXI(2): 259-270

Shavell S (1984) Liability for Harm Versus Regulation of Safety, *The Journal of Legal Studies*, XIII(2): 357-375

Shavell S (2000) On the Social Function and the Regulation of Liability Insurance, Harvard Law School, Law-Econ Discussion Paper No. 278, *Geneva Papers on Risk and Insurance Theory*, http://ssrn.com/abstract=224945 [23.2.2006]

Shirey R (2000) *Internet Security Glossary*, Internet Engineering Task Force (IETF), Request for Comment (RFC) 2828, http://www.ietf.org/rfc/rfc2828.txt [23.8.2005]

Siponen M (2001) An Analysis of the Recent IS Security Development Approaches: Descriptive and Prescriptive Implications, in Dhillon G (ed.) *Information Security Management: Global Challenges in the New Millenium*, Idea Group, London, UK, p. 101-125

Siponen M (2002) *Designing Secure Information Systems and Software. Critical evaluation of the existing approaches and a new paradigm*, Oulu University Press, Oulu, Finland

Siponen MT and Oinas-Kukkonen H (2002) *A Survey of Information Systems Security Issues and Respective Research Contributions*, published as a part of Siponen M (2002) *Designing Secure Information Systems and Software. Critical evaluation of the existing approaches and a new paradigm*, Oulu, Oulu University Press, original paper number I

Skogh G (1977) *Priser, skadestånd och straff*, LiberLäromedel, Lund

Slovic P (2000) Perceived Risk, Trust and Democracy, in Slovic P (2000) *The Perception of Risk*, Earthscan Publications, London, p. 316-327. Reprinted from Slovic P (1993) Perceived Risk, Trust and Democracy, *Risk Analysis*, 13(6): 675-682

Slovic P (2000) Perception of Risk, in Slovic P (2000) *The Perception of Risk*, Earthscan Publications, London, p. 220-232.

Smith DK (2002) What is Regulation? A Reply to Julia Black's 'Critical Reflections on Regulation', *Australian Journal of Legal Philosophy*, 27: 37-47

Smith PG (1999) From Experience: Reaping Benefit from Speed to Market, *Journal of Product Innovation Management*, 16(May): 222-231, also available at http://www.newproductdynamics.com/JPIM5-99/JPIM5-99.pdf [21.2.2006]

vonSolms B and Marais E (2004) From Secure Wired Networks to Secure Wireless Networks – What Are the Extra Risks? *Computers & Security*, 23(8): 633-637

Sommer JH (2000) Against Cyberlaw, *Berkley Technology Law Journal*, 15(3)
http://www.law.berkeley.edu/journals/btlj/articles/vol15/sommer/sommer.html [9.3.2006]

Soo Hoo KJ, Grove GD, Drozdova E, Lukasik SJ, Elliott D and Goodman SE (1999) Regional Interest Group on Information Security: Sharing Information and Exploring Collaborative Opportunities, December 7, 1998, Stanford University, *CISAC Workshop Report*, Center for International Security and Cooperation (CISAC), Stanford,
http://cisac.stanford.edu/publications/10357/ [23.2.2006]

Soo Hoo K, Sudbury AW and Jaquith AR (2001) Tangible ROI through Secure Software Engineering, *Secure Business Quarterly*, 1(2): 1-4 http://www.sbq.com/sbq/rosi/sbq_rosi_software_engineering.pdf [23.2.2006]

Stampley D (2005) Privacy Compliance Enforcement, Part I: Weak Application Security Equals Noncompliance, *InformationWeek*, June 20, 2005,
http://informationweek.com/story/showArticle.jhtml?articleID=164900859 [23.2.2006]

Stapleton J (2004) Regulating Torts, in Parker C, Scott C, Lacey N and Braithwaite J (eds) *Regulating Law*, Oxford University Press, Oxford, USA, p. 122-144

Steinmüller W (1993) *Information Technology and Society. Introduction to Applied Informatics*, Wissenschaftliche Buchgesellschaft, Darmstadt. Summaries in English and fulltext German version of the book Informationstechnologie und Gesellschaft. Einführung in die Angewandte Informatik, is available at
http://www.informaticsapplied-textbook.info/ [20.12.2005]

Strömholm S (1988) *Rätt, rättskällor och rättstillämpning. En lärobok i allmän rättslära*, tredje upplagan, Norstedts, Stockholm

Stuurman C (1991) The EC Directive on Product Liability and its Application to Information Technology, in Meijboom AP and Prins C (eds.) *The Law of Information Technology in Europe 1992*, Kluwer, Deventer, p. 191-207

Stuurman C (1989) Product Liability for Software in Europe. A Discussion of the EC Directive of 25 July 1985, in Vandenberghe GPV (ed.) *Advanced Topics of Law and Information Technology,* Kluwer, Deventer, p. 127-149

Sunstein CR (1990) The Functions of Regulatory Statutes, in *After the Rights Revolution: Reconceiving the Regulatory State*, Chapter 2, Cambridge, Massachusetts: Harvard University Press, p. 52, republished in Ogus AI (ed., 2001) *Regulation, Economics, and the Law,* International Libraty of Critical Writings in Economics Series, No. 137, Edward Elgar Publishing, UK, p. 3-35

Syme S and Camp J (2001) *Code as Governance, The Governance of Code*, Faculty Research Working Papers Series, John F. Kennedy School of Government, Harvard University, No. RWP01-014, April, http://ksgnotes1.harvard.edu/Research/wpaper.nsf/rwp/RWP01-014?OpenDocument [23.2.2006]

Takanen A, Vuorijärvi P, Laakso M and Röning J (2004) Agents of Responsibility in Software Vulnerability Process, *Ethics and Information Technology*, 6: 93-110

Takki P (1998) Vuosi 2000 –tietojärjestelmät ja vastuut yrityksen näkökulmasta, *Defensor Legis*, p. 34-43

Tala J (2005) Lainsäädännön vaihtoehdot – tarve ja tehtävät, in Lindfors H (ed.) *Lainsäädäntöä vai muuta oikeudellista ohjailua*, National Research Institute of Legal Policy, Research Communications, nro. 67, Helsinki, 2005, available at http://www.om.fi/optula/34524.htm [9.3.2006]

Tala J (2005) *Lakien laadinta ja vaikutukset*, Edita, Helsinki

Tala J (2001) *Lakien vaikutukset. Lakiuudistusten tavoitteet ja niiden toteutuminen lainsäädäntöteoreettisessa tarkastelussa* [The Effects of Legislation. Objectives of a Law Reform and their Realization from the Perspective of the Theory of Legislation], Publication of the National Research Institute of Legal Policy, no. 177, Helsinki (with an English summary)

Tamanaha B (2001) *A General Jurisprudence of Law and Society*, Oxford University Press, New York

Tamanaha B (1997) *Realistic Socio-Legal Theory: Pragmatism and A Social Theory of Law*, Clarendon, Oxford

Tapp JL and Levine FJ (1980) Legal Socialization, in Evan WM (ed.) *The Sociology of Law. A Social-Structural Perspective*, The Free Press, New York, p. 121-134. Excerpt from Tapp JL and Levine FJ (1974) Legal Socialization: Strategies ofr and Ethical Legality, *Stanford Law Review*, 27: 1-72

Telang R and Wattal S (2005) Impact of Software Vulnerability Announcements on the Market Value of Software Vendors – an Empirical Investigation, paper presented in the *Fourth Workshop on the Economics of Information Security*, Kennedy School of Government, Harvard University, 2-3 June 2005, available at http://infosecon.net/workshop/pdf/telang_wattal.pdf [9.2.2006]

Teubner G (1986) After Legal Instrumentalism? Strategic Models of Post-Regulatory Law, in Teubner G (ed.) *Dilemmas of Law in the Welfare State*, Walter de Gruyter, Berlin, p. 299-325

Torvund O (1997) *Kontraktsregulering – IT-kontrakter*, Tano Aschehoug, Oslo

Trompenaars B and Hugenholtz PB (1998) *Formation and validity of On-Line Contracts*, Institute for Information Law, Amsterdam, report for the Imprimatur Consortium. Previously available at http://www.imprimatur.net/legal.htm, now on file with the author.

Tuori K (2002) *Kriittinen oikeuspositivismi*, WSLT, Juva

Tuori K (1997) Law, Power and Critique, in Tuori K, Bankowski Z and Uusitalo J (eds.) *Law and Power. Critical and Socio-Legal Essays*, Deborah Charles Publications, Liverpool, p. 7-29

Tuori K (1990) *Oikeus, valta ja demokratia*, Lakimiesliiton kustannus, Mänttä

Tuori K (2003) Tuomarivaltio – uhka vai myytti? *Lakimies, the Journal of the Finnish Lawyers' Association*, 101(6): 915-943

Turunen A (2005) *Innovations as Communication Processes: A Legal Architecture for Governing Ideas in Business*, Lapland University Press, Rovaniemi

Twining W (2003) A Post-Westphalian Conception of Law, *Law & Society Review*, 37(1): 199-257

Ulkuniemi P (2003) *Purchasing software components at the dawn of market*, Acta Universitatis Ouluensis, Oeconomica, G 13, Oulu University Press, Oulu, available at http://herkules.oulu.fi/isbn9514272188/isbn9514272188.pdf [23.2.2006]

Varian H (1996) *Intermediate Microeconomics. A Modern Approach*, 4th edition, W.W. Norton & Company, New York

Varian H (2002) *System Reliability and Free Riding*, paper presented at the Workshop on Economics and Information Security, University of California, Berkley, May 16-17, 2002, http://www.sims.berkeley.edu/resources/affiliates/workshops/econsecurity/econws/49.pdf [23.2.2006]

Vedung E (1998) Policy Instruments: Typologies and Theories, in Bemelmans-Videc M-L, Rist RC and Vedung E (eds.) *Carrots, Sticks & Sermons. Policy Instruments & Their Evaluation*, Transaction Publishers, New Brunswick, USA, p. 21-59

Vedung E (2000) *Public Policy and Program Evaluation*, Transaction Publishers, New Brunswick

Vedung E and van der Doelen FCJ (1998) The Sermon: Information Programs in the Public Policy Process – Choice, Effects, and Evaluation, in Bemelmans-Videc M-L, Rist RC and Vedung E (eds.) *Carrots, Sticks & Sermons. Policy Instruments & Their Evaluation*, Transaction Publishers, New Brunswick, USA, p. 103-129

Viega J and McGraw G (2002) *Building Secure Software: How to Avoid Security Problems the Right Way*, Addison-Wesley, Boston, USA

Viljanen M (2005) Ihmisen identiteetti ja tuottamusarviointi, *Lakimies, the Journal of the Finnish Lawyers' Association*, 103(3): 426-452

Viljanen V-P (2001) *Perusoikeuksien rajoitusedellytykset*, WSLT, Helsinki

Viscusi WK and Moore MJ (1993) Product Liability, Research and Development, and Innovation, *The Journal of Political Economy*, 101(1): 161-185

Voas JM (1998) Certifying Off-the-Shelf Software Components, *IEEE Computer*, 31(6): 53-60

Välimäki M (2005) *The Rise of Open Source Licensing: A Challenge to the Use of Intellectual Property in the Software Industry*, Turre Publishing, Helsinki

Väntsi R (1994) *"Varoitanko kuluttajaa tuotteen vaaroista?" Tuotevastuu ja tuoteturvallisuuslainsäädäntö elinkeinonharjoittajan informointipäätösten perusteena*, Acta Universitatis Tamperensis, Series A 431, University of Tampere, Tampere

Wahlgren P (2003) *Juridisk riskanalys. Mot en säkrare juridisk metod*, Jure AB, Stockholm

Warsta J (2001) *Contracting in Software Business. Analysis of Evolving Contract Processes and Relationships*, Oulu University Press, Oulu

Weber M (1949) *The Methodology of the Social Sciences*, Free Press, New York

Webster F (1995) *Theories of the Information Society*, Routledge, London and New York

Weimer DL and Vining AR (1999) *Policy Analysis. Concepts and Practice*, 3rd ed., Prentice Hall, Upper Saddle River, New Jersey

Weiss JA (2002) Public Information, in Salamon LM (ed.) *The Tools of Government. Guide to the New Governance*, Oxford University Press, p. 217-255

Wesselius J and Ververs F (1990) Some elementary questions on software quality control, *Software Engineering Journal*, 5(6): 319-330

Wheterbe JC and Frolick MN (2000) Cycle Time Reduction: Concepts and Case Studies, *Communications of the Association for Information Systems*, 3(13), May, http://cais.isworld.org/articles/default.asp?vol=3&art=13 [23.2.2006]

White MJ (1989) An Empirical Test of the Comparative and Contributory Negligence Rules in Accident Law, *RAND Journal of Economics* 20(3): 308-330. Reprinted in De Geest G and Van den Bergh R (eds., 2004) Comparative Law and Economics Volume II, The International Library of Critical Writings in Economics 170, Edward Elgar Publishing Limited, Cheltenham, UK, p. 191-213

Wikström K (1998) Ohjaaminen oikeusnormeilla, *Juhlajulkaisu Leena Kartio 1938 – 30/8 – 1998*, Turun yliopiston oikeustieteellisen tiedekunnan julkaisuja A. Juhlajulkaisut N:o 7, Turun yliopisto, Turku

Wikström K (1994) *Oikeus ja talous. Tutkimus markkinamekanismien normatiivisen ohjaamisen tarpeesta ja mahdollisuuksista – esimerkkinä verotus*, Lakimiesliiton kustannus, Helsinki

Wilhelmsson T (1996a) Administrative Procedures for the Control of Marketing Practices – Theoretical Rationale and Perspectives, in Junkkari T (ed.) *Twelve Essays on Consumer Law and Policy*, Publications of the Department of Private Law, University of Helsinki, Yliopistopaino, Helsinki, p. 142-164. Originally published in *Journal of Consumer Policy*, Vol. 15, 1992, p. 159-177. References are made to the reprinted version

Wilhelmsson T (1996b) Contribution to a Green Sales Law, in Junkkari T (ed.) *Twelve Essays on Consumer Law and Policy*, Publications of the Department of Private Law, University of Helsinki, Yliopistopaino, Helsinki, p. 267-287. Originally published in Swedish "Bidgra till en grön köprätt", in Blume P and Petersen H (eds. 1993) Retlig Polycentri, Copenhagen, Akademisk Forlag, p. 19-36

Wilhelmsson T (2000) Korvausvastuu uutena sääntelyvälineenä, in Wilhelmsson T et al., *Pieniä kertomuksia hyvinvointivaltion siviilioikeudesta*, WSLT, Helsinki

Wilhelmsson T (1994) Köprätten och produktansvaret, *JFT* 1994(6): 627-642

Wilhelmsson T (2001) *Senmodern ansvarsrätt. Privaträtt som redskap för mikropolitik*, [Late Modern Liability Law, Private Law as a Tool for Micropolitics], Kauppakaari, Helsinki

Wilhelmsson T (1995a) *Social Contract Law and European Integration*, Darthmouth, Aldershot, England

Wilhelmsson T (2003) Sopimusoikeuden eurooppalaistuminen ja oikeudellisten toimijoiden roolit – useampi kokki keittää paremman sopan [The Europeanisation of Contract Law and the Roles of Various Legal Actors], *Lakimies, the Journal of the Finnish Lawyers' Association*, 101(7-8): 1098-1119

Wilhelmsson T (1991) *Suomen kuluttajansuojajärjestelmä*, Lakimiesliiton kustannus, Helsinki

Wilhelmsson T (1995b) *Vakiosopimus. Sopimussidonnaisuudesta ja kohtuuttomista sopimusehdoista*, Lakimiesliiton Kustannus, Helsinki

Wilhelmsson T and Rudanko M (2004) *Tuotevastuu*, 2. uudistettu painos, Talentum, Helsinki

Wintgens LJ (2002) Legislation as an Object of Study of Legal Theory: Legisprudence, in Wintgens LJ (ed) *Legisprudence: A New Theoretical Approach to Legislation*, Hart Publishing, Oxford, p. 9-41

Wintgens LJ (2005) Legisprudence as a New Theory of Legislation, in Wintgens LJ (ed) *The Theory and Practice of Legislation: Essays in Legisprudence*, Ashgate, Aldershot, p. 3-26

Ylikortes K (1992) *Tuotevastuu. Tuotevastuulain vaikutukset PKT-yrityksiin*, Publication of the Helsinki Research Institute for Business Administration, Series B 80, Helsinki

**Official Sources**

*OECD*

OECD (2003) *Government Capacity to Assure High Quality Regulation in Finland*, OECD Reviews of Regulatory Reform, Paris, http://www.oecd.org/dataoecd/32/52/2510133.pdf [23.2.2006]

OECD (2002) *Guidelines for the Security of Information Systems and Networks*, Recommendation of the OECD Council at its 1037th Session on 25 July 2002, OECD, available at http://www.oecd.org/document/42/0,2340,en_2649_34255_15582250_1_1_1_1,00.html [23.2.2006]

OECD (1997) *Regulatory Impact Analysis: Best Practices in OECD Countries*, OECD Publications, Paris

OECD (2002) *Regulatory Policies in OECD Countries. From Interventionism to Regulatory Governance*, OECD Reviews of Regulatory Reform, OECD, Paris

OECD (2004) *Summary of responses to the survey on the implementation of the OECD guidelines for the security of information systems and networks: Towards a culture of security*, DSTI/ICCP/REG(2003)8/FINAL, Paris: Directorate for Science, Technology and Industry; Committee for Information, Computer and Communications Policy; Working Party on Security and Privacy, available at http://www.olis.oecd.org/olis/2003doc.nsf/LinkTo/dsti-iccp-reg(2003)8-final [23.2.2006]

OECD (2005) *The Promotion of a Culture of Security for Information Systems and Networks in OECD Countries*, DSTI/ICCP/ REG(2005)1/ FINAL, Paris: Directorate for Science, Technology and Industry; Committee for Information, Computer and Communications Policy; Working Party on Security and Privacy, http://www.oecd.org/dataoecd/16/27/35884541.pdf [25.4.2006]

*EU*

COM(1995)617 final, First Report from the Commission on the Application of Council Directive on the Approximation of Laws, Regulations and Administrative Provisions of the Member States Concerning Liability for Defective Products (85/374/EEC), Brussels, 13.12.1995, http://europa.eu.int/comm/enterprise/regulation/goods/docs/liability/ com-95-617/ com-95-617_en.pdf [23.2.2006]

COM(1999)396 final, Commission Green Paper on Liability for Defective Products, Brussels 28th July 1999, available at http://europa.eu.int/comm/enterprise/regulation/goods/docs/liability/ 1999-greenpaper/ com1999-396_en.pdf [15.2.2006]

COM(2000)199 final, Report from the Commission to the Council, the European Parliament and the Economic and Social Committee on the implementation of and effects of Directive 91/250/EEC on legal protection of computer programs, 10.4.2000

COM(2000)248 final, Report from the Commission on the Implementation of Council Directive 93/13/EEC of 5 April 1993 on Unfair Terms in Consumer Contracts, Brussels, 27.04.2000, http://europa.eu.int/comm/consumers/cons_int/safe_shop/unf_cont_terms/uct03_en.pdf [23.2.2006]

COM(2000)890, Communication from the Commission to the Council, the European Parliament, The Economic and Social Committee and the Committee of Regions, Creating a Safer Information Society by Improving the Security of Information Infrastructures and Combating Computer-related Crime

COM(2000)893 final, Report from the Commission on the Application of the Directive 85/374 on Liability for Defective Products, Brussels, 31.1.2001, http://europa.eu.int/eur-lex/lex/LexUriServ/site/en/com/2000/com2000_0893en01.pdf [12.8.2005]

Communication from the Commission to the Council and the European Parliament on European contract law, Official Journal C 255 , 13/09/2001 P. 0001 – 0044

COM(2001)298 final, Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of Regions of 6.6.2001, Network and Information Security: Proposal for a European Policy Approach

COM(2003)68 final, Communication from the Commission to the European Parliament and the Council - A more coherent European contract law - An action plan, Official Journal C 063, 15.03.2003, p. 1 – 44

Opinion of the European Economic and Social Committee on the Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions on Network and Information Security: Proposal for a European Policy Approach, COM(2001) 298 final, TEN/083, Brussels, November 28, 2001, available at the home pages of the EESC at http://www.esc.eu.int/index_en.asp [28.12.2005]

*European Court of Justice (ECJ)*

Mariá Victoria Gonzáles Sánchez v Medicina Asturiana SA, Case C-183/00 [2002]

Commission v France, Case C-52/00 [2002]

Commission v Greece, Case C-154/00 [2002]

*Finland*

Government Proposal 231/2005, Hallituksen esitys Eduskunnalle viestintämarkkinalain ja eräiden markkinaoikeudellisten asioiden käsittelystä annetun lain muuttamisesta

Ministry of Transport and Communications Finland (MINTC, 2001) *Kansallisen tietoturvastrategian tarve Suomessa* [Does Finland need a national information security strategy?], Selvitys – Report, available at http://www.mintc.fi/www/sivut/dokumentit/julkaisu/mietinnot/2001/36.htm [26.4.2003] (in Finnish with English description)

Ministry of Transport and Communications Finland (2004) *Turvalliseen tietoyhteiskuntaan: Kansallisen tietoturvallisuusasioiden neuvottelukunnan toimintasuunnitelma*, Ohjelmia ja strategioita 1/2004, Liikenne- ja viestintäministeriö, Helsinki, available at http://www.mintc.fi/oliver/upl719-Toimintasuunnitelma.pdf [20.3.2006]

Secretariat of the National Information Security Advisory Board (2004) *Creating a Safer Information Society. National Information Security Advisory Board report submitted to the Government on 14 December 2004*, Programmes and strategies 3/2005, Ministry of Transport and Communications, Helsinki, available at http://www.mintc.fi/oliver/upl499-OS_3_2005%20(midres).pdf [20.3.2006]. Originally published in Finnish as *Tietoturvalliseen tietoyhteiskuntaan. Kansallisen tietoturvallisuusasioiden neuvottelukunnan kertomus valtioneuvostolle 14.12.2004*, Ohjelmia ja strategioita 1/2004, available at http://www.mintc.fi/oliver/upl165-OS%204_2004.pdf [20.3.2006]

Kansallisen tietoturvallisuusasioiden neuvottelukunnan sihteeristö [Secretariat of the National Information Security Advisory Board] (2006) *Tietoturvalliseen tietoyhteiskuntaan. Kansallisen tietoturvallisuusasioiden neuvottelukunnan kertomus valtioneuvostolle 13.12.2005* [Creating a Safer Information Society. National Information Security Advisory Board report submitted to the Government on 13 December 2005], Liikenne- ja viestintäministeriön julkaisuja 93/2005, Liikenne- ja viestintäministeriö (in Finnish, with an English description), available at http://www.mintc.fi/oliver/upl179-Neuvottelukunnan%20raportti.pdf [20.3.2006]

VYSE 1998, Valtion tietotekniikkahankintojen yleiset sopimusehdot 1998, http://www.vm.fi/tiedostot/pdf/fi/16877.pdf [9.1.2006]

Valtioneuvosto (2004) *Hallituksen esityksen laatimisohjeet*, Oikeusministeriön julkaisuja 2004:4, Helsinki

Valtioneuvoston lainvalmistelun suunnittelun ja johtamisen kehittämisryhmän mietintö (Lainvalmistelun kansliapäällikköryhmä), *Tehokkaampaa, suunnitelmallisempaa ja hallitumpaa lainvalmistelua*, Valtioneuvoston kanslian julkaisusarja 13/2005, Edita 2005, available at http://www.vnk.fi/vn/liston/vnk.lsp?r=98044&k=fi&old=954 [25.11.2005] (in Finnish)

Valtionhallinnon tietoturvallisuuden johtoryhmä VAHTI (2001) Valtion tietotekniikkahankintojen tietoturvallisuuden tarkistuslista, Valtiovarainministeriö, Ohje VAHTI 6/2001, available at http://www.vm.fi/vahti [9.1.2006]

*USA*

Chambers JT and Thompson JW (2004) *Vulnerability Disclosure Framework*, Final Report and Recommendations by the Council, National Infrastructure Advisory Council (NIAC), Vulnerability Disclosure Working Group, http://www.dhs.gov/interweb/assetlibrary/vdwgreport.pdf [11.2.2006]

Computer Science and Telecommunications Board, System Security Study Committee (1991) *Computers at Risk. Safe Computing in the Information Age*, National Academy Press, Washington D.C., http://books.nap.edu/books/0309043883/html/index.html [30.8.2005]

Computer Science and Telecommunications Board CSTB (2002) *Cybersecurity Today and Tomorrow: Pay Now or Pay Later*, National Academy Press, Washington D.C., http://books.nap.edu/html/cybersecurity/ [23.6.2004]

NCSP National Cyber Security Partnership (2004) *Improving Security Across the Software Development LifeCycle*, Security Across the Software Development Life Cycle Task Force Report, April, http://www.cyberpartnership.org/SDLCFULL.pdf [13.1.2006]

NCSP National Cyber Security Partnership (2004) *Technical Standards and Common Criteria*, Technical Standards and Common Criteria Task Force, Report, http://www.cyberpartnership.org/TF4TechReport.pdf [13.1.2006]

Schneider FB (ed., 1999) *Trust in Cyberspace*, Committee on Information Systems Trustworthiness, Computer Science and Telecommunications Board, National Research Council, National Academy Press, Washington, D.C., http://bob.nap.edu/html/trust/ [15.2.2006]

*Web Pages*

CERT® FAQ (2005) Frequently Asked Questions about CERT, at http://www.cert.org/faq/cert_faq.html [updated 15.12.2005, visited 23.1.2006]

Computer Security Incidence Response Team (CSIRT) Frequently Asked Questions (FAQ) (2002) at http://www.cert.org/csirts/csirt_faq.html [updated 1.2.2002, visited 23.1.2006]

International Telecommunications Union, Telecommunication Standardization Sector (ITU-T), Study Group 17, ICT Security Standards Roadmap v1.0, November 2005, at http://www.itu.int/ITU-T/studygroups/com17/ict/index.html [updated 25.1.2006, visited 1.2.2006]

# List of abbreviations

| | |
|---|---|
| BSA | Business Software Alliance |
| CAC (C&C) | Command and control |
| CPSR | Computer Professionals for Social Responsibility |
| CSIRT | Computer Security Incidence Response Team |
| COTS | Commercial Of-the-Self |
| DDoS | Distributed Denial of Service |
| DL | Defensor Legis |
| DMCA | Digital Millennium Copyright Act |
| ECJ | European Court of Justice |
| ENISA | European Network and Information Security Agency |
| EPIC | Electronic Privacy Information Center |
| EU | European Union |
| EULA | End-User License Agreement |
| FICORA | Finnish Communications Regulatory Authority |
| IA | Internet Alliance |
| IEC | International Electrotechnical Commission |
| IETF | Internet Engineering Task Force |
| ICT | Information and Communication Technology |
| IFIP | International Federation for Information Processing |
| IPR | Intellectual Property Right |
| IS | Information System |
| ISO | International Organization for Standardization |
| IT | Information Technology |
| ITSEC | Information Technology Security Evaluation Criteria |
| JFT | Tidskrift utgiven av Juridiska Föreningen I Finland |
| JTC | Joint Technical Committee of ISO and IEC |
| NGO | Non-Governmental Organisation |
| NSA | Non-State Actor |
| NIAC | National Infrastructure Advisory Council of the U.S. |
| OIS | Organisation for Internet Safety |
| OJ | Official Journal of the European Union |

| | |
|---|---|
| OUSPG | Oulu University Secure Programming Group |
| OWASP | Open Web Application Security Project |
| PC | Personal Computer |
| RFC | Request for Comments |
| SME | Small or Medium-Sized Enterprise |
| SW | Software |
| TCSEC | Trusted Computing System Evaluation Criteria |
| TSG | Trusted Computing Group |
| W3C | World Wide Web Consortium |

# 1 Introduction

Henrik Kaspersen has stated, while considering the possibilities to regulate information security 'already' in 1993, that "information security should be an integrating part and a standing quality factor of information products and systems, without which IS [i.e., information system: authors note] could not maintain its predominant place in our post-modern organisational structures. However, what exactly should be enhanced? How should it be enhanced and what factors explicitly should be taken into consideration? Where does the law come in and what can it do?"[1]. He was knowledgeable enough to recognise that the discussion was still too young and needs to find its direction. No adequate overall solutions to the problems of information security could be provided at the time.

This actually is still true, even though over ten years have passed. Solutions for the problems have been provided, but the hard problem of insecurity of information systems is still unanswered. It is not expected to be resolved any time soon, at least not by a silver bullet or any small number of bullets. What can be done, and what is increasingly being done, is to find solutions to the smaller problems inside the overall information security dilemma.

Contributions from the technological and social side have been made on several subsections such as the control of access to information, secure communication, management of information security, and development of secure information systems and

---

[1] Kaspersen, Foreword, p. 8. Henrik Kaspersen is currently Director of the Computer Law Institute at the Faculty of Law of the Free University of Amsterdam in Netherlands (http://cli.vu/en/wiecli.php) [6.3.2006]. He is probably best known for his work as a Chairman of the Committee of Experts on Crime in Cyberspace of the Council of Europe that prepared the Convention on Cybercrime (CETS nro: 185) that entered into force 1.7.2004.

software[2]. However, much of current information security discussion (both among researchers and practitioners) is about access control and secure communication – about detecting, preventing, and responding to attacks and vulnerabilities, especially in networks.

But all of this effort addresses symptoms of a problem that has been severely neglected for long; the lack of secure software and information systems development. A vast majority of attacks make use of software vulnerabilities that are entirely preventable and take advantage of the lack of security properties in systems. As stated by a pair of leading experts in secure software development, John Viega and Gary McGraw, in their textbook on secure software development "behind every computer security problem and malicious attack lies a common enemy – bad software"[3]. The most frequently exploited vulnerabilities are defects in common software products. For information security to be able to live up to the expectations, the security of the underlying software and information systems needs to be improved[4].

---

[2] According to the classification of the subject matters of the young discipline of information security made by Siponen and Oinas-Kukkonen in *A Survey of Information Systems Security Issues and Respective Research Contributions*, p. 4-5, *access control* includes issues like access control models and policies, information flow control models, operating systems protection, anti-virus techniques, water-marking, and cognitive passwords. *Secure communication* includes issues from cryptographical algorithms and systems, virtual private networks (VPN), anonymity techniques and stenography. *Information security management* issues concern issues on firewalls, digital signatures, auditing, intrusion detection systems, and security policies. The issue of *development of secure information systems and software* includes secure programming, methods for developing and managing secure information systems and software, risk analysis, and testing methods.

[3] Viega and McGraw, *Building Secure Software*, p. 1.

[4] The improvement of the security of software and information systems is not the only thing needed to secure the information infrastructure, but it is increasingly being recognised as an important element that has been largely neglected.

Only quite recently has the issue of developing secure software started to gain attention[5]. For example, the OECD report that only few initiatives regarding the development of secure software has been initiated by its member states by the end of 2005[6].

It is no surprise that the same goes with the regulation concerning information security. There is a large and growing amount of regulation (together with legal problems still in need to be solved and others corrected) concerning mainly access control, secure communication and information security management issues like criminalisation of unauthorised access, regulation of cryptography, anonymity, digital signatures, data protection (as protection of privacy), computer evidence, confidential communications, etc[7]. Issues concerning secure software and information system development have largely been neglected as well in the regulatory area (especially by governmental regulators) as by legal scholarship[8].

---

[5] Secure software development has not been, and still largely is not, part of the curriculum in most educational institutions teaching software developers. Not even in the country producing most software, i.e., USA, as made explicit in the report of a task force of the National Cyber Security Partnership, *Improving Security Across the Software Development LifeCycle*, p. 3-4, concerning year 2004. Even textbooks on secure coding and design for software engineers or information system developers have been lacking for a long time, at least in English language. In Finland, the issue of secure software development has largely been missing in the university curriculum, despite the relatively long tradition of research and teaching in information security. See, e.g., the report of state of research and teaching of information security in Finland in October 2004 conducted by Marko Helenius, *Tietoturvallisuuden tutkimus ja opetus*.

[6] See the OECD report, *The Promotion of a Culture of Security for Information Systems and Networks in OECD Countries*, p. 17 and 22, analysing governments' effective efforts to foster a shift in security culture.

[7] Regulation concentrates largely on computer crime issues and on the information security responsibilities of those processing personal data or providing electronic communication networks and services. There is also relatively extensive regulation on the information security obligations of governmental agencies. For examples, see chapter 5 below.

[8] Perhaps it is due to the close links to the science and art of information security that legal informatics has largely concentrated on the same issues as mainstream information security research. Note that information security as an

As an effort to bridge this gap the study concentrates on the regulation of secure software and information systems development of the different disciplines of information security. The general argument, and the fundamental position adopted, is that also regulation should address the reasons for the insecurity rather than just to react to the symptoms. In this study, following Kaspersen's argument, information security is considered to be an important societal objective - a necessity in the networked society and especially its information infrastructure without which we, as citizens, could not use our fundamental rights in the networks[9]. As an answer to

---

object of study in legal scholarship is by no means new even in relation to ICT. As the German founder of legal informatics as a discipline, Wilhelm Steinmüller, notes in presenting legal informatics as an example of the way applied informatics works, legal informatics have had to propose solutions for urgent social and legal issues of computer applications such as data security and privacy from its beginning in 1969. See Steinmüller, *Information Technology and Society*, English abstract of Book One – Science Theory of Applied Informatics. Peter Seipel, another pioneer of the discipline of legal informatics and one of its major figures especially in the Nordic countries, saw already in 1977 that information security is one of the crucial interfaces of ICT and law, and even an independent field of legal study that calls for an integrated approach bringing together all branches of law, in his still up-to-date doctoral dissertation, *Computing Law*, p. 160 and 262. The development of legal informatics and its interest in information security is depicted by Ahti Saarenpää in Tietojenkäsittelystä läsnä-älyyn – katkelmia oikeusinformatiikan kehityksestä, p. 91-123.

[9] In this sense, information security is seen as a collective good, a guarantee of the functioning of the network society. Of this discussion in Finland see, e.g., the most recent contribution of Tuomas Pöysti in ICT and Legal Principles, p. 560-600. However, information security can also be seen as part of our fundamental right to security at the individual level – as an individual right instead of a collective good. At this individual level information security protects our identity and our right to informational self-determination. Of this discussion in Finland, see the report of the Institute for Law and Informatics to the Ministry of Transport and Communications Finland, Saarenpää et al., *Sähköinen viestintä, tietoturvallisuus ja perusoikeudet*. The report is part of coordination of the actions implementing Finland's National Information Security Strategy that has been the task of the National Information Security Advisory Board since it began its work in spring 2004. More information about Finland's National Information Security Strategy and of the work of the

Henrik Kaspersen's original question, I posit that secure software development is a crucial element in improving information security. It is that kind of a conduct that would be desirable for the law and regulation in general to promote or encourage, to give some motive or ground for or, at least, not to hamper unnecessarily[10].

This does not mean that the work done in order to minimise the effects of these symptoms is useless. In contrary, it is very well needed since the reasons for the vulnerabilities are not going to be altered in any near future. There has been a strong need, which is not vanishing, for corrective and preventive information security measures. However, I think that it is time to start caring for the causes more seriously also in the regulatory arena and not just to assume that vulnerabilities in the current magnitude are something that just has to be lived with.

Before we can continue, there is a need to specify the problem area and show some central terms. In this chapter the concept used are clarified. Especially the issues involved with the use of concepts information security and regulation are considered. At the same time

---

Advisory Board, see the web pages of the Ministry of Transport and Communications Finland, at http://www.mintc.fi/ (at the English language portal see Communications – Information Security and Privacy Protection – National Information Security Strategy) [16.3.2006]

[10] Thus, secure software development is seen as an important societal objective or a public policy in the sense Neil MacCormick, On Legal Decisions and their Consequences, p. 255 footnote 39, understands it in considering the different grounds or criteria used by judges in the evaluation of juridical consequences and possible ulterior outcomes of possible rulings. Similar approach to information security in general has already been adopted in the guidelines for the drafting of government proposals in Finland (Valtioneuvosto, *Hallituksen esityksen laatimisohjeet*), where one of the social effects of the law that have to be considered when drafting a parliamentary act are the effects on the information society. The possible effects on information security have to be considered as part of them. As a statement of the importance of information security this is a good development as such, even though the drafting guidelines are not widely used in practice as a recent report on the state and development of law-drafting in Finland shows (Lainvalmistelun kansliapäällikköryhmä, *Tehokkaampaa, suunnitelmallisempaa ja hallitumpaa lainvalmistelua*, p. 211-212).

the object of research is depicted and qualifications made. At the end of this chapter the purpose of the study and the research questions are made explicit together with the methods used and the central limitations of the study.

## 1.1 Seeing through a conceptual muddle

Information security is still in that phase of research where almost every study starts with definitions of the terms used. And these definitions and the terms used seem to vary even from researcher and research area to another. Not to even mention the terms used and definitions given by practitioners. With just a little exaggeration it can be said that the researchers in different areas of information security belonging to different paradigms are not aware of what others are saying and doing and may not even recognise the alternative views of reality which lie outside their boundaries[11]. And not even the professional field is using the same terms due to the specialisation into different areas of information security work.

Due the diversity of approaches and lack of generally accepted use of terms information security is in a stage of conceptual muddle. Actually there is not even an agreement on which term to use and what areas they cover. Terms such as computer security, IT security, information system security, information security, and data security are, among others, used rather synonymously with an effort to depict the whole profession and the research area with variable subsections.

---

[11] As Mikko Siponen, *Designing Secure Information Systems and Software*, p. 45, point out in his doctoral dissertation, information security research and practice is done on several disciplines that stem from separate scientific backgrounds; information system security is rooted in information system science, computer and database security in computer science, cryptology and communications security in mathematics, and practitioners usually come from the software engineering side. Since these research areas stem from different paradigms in the sense of Burrel and Morgan, *Sociological Paradigms and Organisational Analysis*, p. 23, and thus provide different perspectives, they also generate quite different concepts and analytical tools.

Of the more recent terms information assurance and security of information infrastructure are used mainly by the military and government agencies in the context of national security and together with information warfare or information operations. This diversity in approaches and terms used creates real difficulties also for lawyers and especially legislators in piecing the whole area.

These afore mentioned wide terms are used to piece together the different areas of specialisation in information security or to highlight parts of them: computer security, communications security, database security, hardware security, software security, personnel security, cryptology, network security, information system security management, organisational level information system security, information security awareness, etc. The definitions of these different areas vary and they have partly overlapping coverage.

In this research terms information security and information system security are used as generic upper level concepts; they depict the whole area of research and the profession as a totality. These terms are often used almost synonymously. However, the difference between the terms information and information systems should be acknowledged; information and information systems are separate but interdependent entities in the same way as products and production processes. Information system security is thus about the security of the different components of the system (hardware, software, infrastructures, users etc.) and information security defined literally (narrowly) about the security of the object of the process in the system. However, there are several factors where system properties affect the security of content and vice versa, so the separation is not absolute.

The main reason the above usage of terms is that information is the target of protection and information systems the processing environment in the different disciplines (information system security, computer and database security, cryptology and practitioners). Information security also emphasises the vital protected asset (information) of organisations and societies among other assets (e.g. personnel/citizens and property). Another reason for favouring these terms to depict the whole area is that they are wide enough to cover

all relevant issues (administrative and organisational information security, personnel security, physical security, communications security, hardware security, software security, data security and operations security), even though they might be too wide and thus empty in other circumstances[12]. More specific (and more meaningful) terms are used when dealing only with a special area of information security.

Information security is not particularly well defined even in the different areas of research (information system, computer science, software engineering, mathematics) or among practitioners and thus prone to adjustments. It is a concept that transforms while the society evolves, because it depicts the relationship between different relational phenomena. Many context dependent definitions of the term are inadequate under a different set of assumptions. This is why no formal definition of the general term of information security is provided in this study[13]. It is assumed to be known at the general level and more detailed conceptualisations are not needed for the research interests relevant to this study. The concentration is just on one of the research and practice areas, i.e., the development of secure software, and not on information security in general. A formal definition is neither needed for qualificatory purposes since the qualifications are made by analysing the subject matter further, and not by formalising

---

[12] Especially information security as a term depicting the whole profession and research area is too wide and thus meaningless for many purposes. Its ability to classify the issue of information security is pretty low and it does not ease the understanding of it.

[13] Information security in general does not need to be defined explicitly to cover every possible situation. A formal definition applicable to all situations of it is as much needed as definitions of terms like contract or privacy. The threats of formal definitions of the term privacy in specific parliamentary acts to the understanding of the role of privacy in society and the difficulties in providing a generally applicable definitions has been noted by Ahti Saarenpää in Yksityisyys, yksityiselämä, yksilön suoja, p. 313-337, in analysing the conceptual environment surrounding the term privacy as a relational concept.

## 1.2 Qualification of modes of software and information system

In this study the concentration is on software development, not the development of information systems as such. Even though an absolute separation of them is not possible software can be seen as a component of an information system. Software often is a central element, though a system almost always includes a number of software products and other elements. However, computer based information system development is dragged along to the degree it compares to software development.

> Two different kinds of information systems of an organisation can be separated; the formal and the informal. The boundary between them is made on the basis of which routinely handled operations can be formalised. A computer based information system, an automation of a part of the formal information system of an organisation, includes the software, hardware (for processing and communication), people, and rules that make a collection of software products work for the user.[14]
>
> Note that the background assumptions and research approaches in information system and software engineering communities differ. While the former community takes into account also social and organisational aspects and uses a variety of research approaches (also those common to social sciences), the latter traditionally focuses on practical means of developing software[15]. However, both fields of

---

[14] According to Gurpreet Dhillon, *Managing Information System Security*, p. 3-5, a computer based information systems can be used to automate only a small part of the formal system.

[15] This separation between the fields of research has also been made in the context of secure information system and software development, for example, by Mikko Siponen in his doctoral dissertation *Designing Secure Information Systems and Software*, p. 46. Notice that this separation is not always necessary – it depends on the approach used and the angle needed in the study in question. See, e.g., the research of agile software development methods by Abrahamsson et al., in *Agile Software Development Methods: Review and Analysis*, p. 7, where this separation was not considered necessary.

> research have faced similar problems in relation to secure development. In both, and especially in the methods used for the development, the security design aspects have been neglected thus requiring separate methods for the development of secure software and information systems[16].

It is not appropriate to speak even of software as a whole – not even in terms of development. Neither information systems nor software are unambiguous and indivisible concepts. There are different software business models that differ also in the ways development is done[17]. Typical separation of software business models is between commercial-off-the-self (COTS), modified-of-the-self (MOTS), tailored and embedded. There are also different options for developing products as a part of business model.

> As an example, following options have been defined. The *core product* option focuses on the development of a single product or product family to be delivered to several customers as is. Development is done prior to sales and customization according to individual user requirements by parameterisation or tailoring is not possible in this option. The *plausible promise* option differs from the core product option in the sense that when the product is first made available to users, it is not finalized in terms of functionality or quality (this is typical in open source development but also in internet software development). The *parameterised product* option focuses on the

---

[16] This has been emphasised by Richard Baskerville at the beginning of the 90's in The Developmental Duality of Information Systems Security, p. 1, and in Information Systems Security Design Methods, p. 375. Also Mikko Siponen emphasises this point in the beginning of the 21st century in An Analysis of the Recent IS Security Development Approaches, p. 101, and in *Designing Secure Information Systems and Software*, p. 13. Similarly also in Devanbu and Stubblebine, Software Engineering for Security, p. 227-228.

[17] This is pointed out, e.g., by Juhani Warsta in *Contracting in Software Business*, p. 34, while analysing the contract processes and relationships in the software business in a comprehensive manner in 2001. According to Rajala et al., Software Business Models, p. 7, the concept of business model is defined as "… an action plan for a company in a given life cycle phase and under certain market conditions".

> development of a customisable product that can be tailored to a degree. In the *product platform* option the focus of development is on a uniform core of several products or customer specific solutions and *project* option focuses on the tailor-made solution to customer's needs even on a one time basis.[18]
> The COTS business model in concentrating on software products that are not customized (tailored) according to individual user requirements corresponds with the core product development option. The MOTS business mode is similar to the parameterised product with the definition of "customisable product that can be tailored to a degree". The tailored approach is equivalent to the project that focuses on the tailor-made solution to customer's needs. But as Juhani Warsta notes, the categorization is a question of definition that is emphasized by the angle of approach used and needed in each separate study.[19]

In the context of considering the development of secure software and information systems the separation between the extremes of COTS and tailored is sufficient or as discussed especially in software engineering; between market-driven (packaged) software development and bespoked (tailor made) software or information system development. The development procedure in these business models is so different that it is useful to separate them.

The characteristic differences between traditional tailor made (bespoked) or information system and market-driven software development are fundamental organisational issues, such as the primary goal (compliance to requirements specification/time-to-market), the success measurements (satisfaction, acceptance/sales, market share, product reviews) and the product life cycle (one release,

---

[18] Rajala et al., Software Business Models, p. 40-43 and p. 51.

[19] Warsta, *Contracting in Software Business*, p. 34. Rajala et al., Software Business Models, p. 41) see the commercial software even going towards a plausible promise option. This will be discussed further in the following section in terms of market-driven development.

then maintenance/several releases, as long as there is a market for the product)[20].

In this study, the concentration is on market-driven (packaged) software development; developing for a market-place (conceivable market segments) rather than for a particular customer or a (group of) user(s) based on a contract. This mode of development is used both in COTS and embedded software business models. Other forms of software or information system development and the related business models are not dealt with, because the market incentives and development practices are so different[21]. For example, the network effects do not play such an important role in them.

This qualification is further justified by the expectedly continuing trend in software development to develop and embrace technologies that reduce the amount of new programming (which is difficult, labour-intensive, and time-consuming), hence to reduce the costs involved in developing any software or information system. Perhaps the most visible example of the trend to avoid programming functionality from scratch, together with in-house software reuse

---

[20] See, e.g., Natt och Dag, *Elicitation and Management of User Requirements in Market-Driven Software Development*, p. 17-18; Carlshamre, *A Usability Perspective on Requirements Engineering*, p. 58-59; Sawyer, A market-based perspective on information systems development, p. 97. These differences will be considered further in the heading where their effects to the security of the software are considered.

[21] A parallel analysis of the effects of regulation on both COTS and open source software (OSS) as business models and development methods is not justified even though it might create useful insights in relation to their secure development. The incentive structures, business methods and their influence on the development are so different that the complexity of the analysis would become overwhelming. Instead, OSS development method could be seen as a process-based regulatory instrument that tries to alter the incentive structures of software development towards better security consciousness. It could be studied as a counterforce, as a shift back to the origins of software development (a time before extensive proprietary software development), that rivals with the market-driven software development characterising COTS business. This discussion has not, however, been taken further in this study. A regulatory analysis of OSS development method is a task for future research.

(reusable software assets developed in-house), is the increased use of (COTS) software components (systems, subsystems, and libraries of components)[22]. The proliferation and falling relative prices for commercial software means that organizations that once would develop systems they wanted themselves are more likely to buy at least components if not entire systems. These components are used even in safety critical systems[23].

Software-based systems are no longer developed in-house and built from the scratch by each user organisation or for them. Instead, many (if not most) of information system and software pieces (of which the system is put together) are bought ready to be installed. Studies in software economics show that the economics of software development leave system designers with no choice but to use large COTS (packaged) components in their systems[24].

---

[22] This has been pointed out, e.g., by Fred B. Schneider in an influential U.S. National Research Council report from 1999, *Trust in Cyberspace*, p. 281, and by Pär Carlshamre in his doctoral dissertation from 2001, *A Usability Perspective on Requirements Engineering*, p. 6. Reasons cited for the increasing use of COTS software components include, in addition to the minimisation of development and maintenance costs, lack of software developers inside organisations, and the desire to reduce risks associated with software development. An important benefit of using COTS components has also been considered the possibility of attaining shorter time-to-market for the final products. The advantages and difficulties in using commercial software packages to improve software productivity has been discussed already in 1981 by Barry Boehm in *Software Engineering Economics*, p. 647-654. This has been pointed out also by Pauliina Ulkuniemi in her doctoral dissertation in economics, *Purchasing software components at the dawn of market*, p. 103, with reference to Feblowitz MD and Greenspan SJ (1998) Scenario-Based Analysis of COTS Acquisition Impacts, *Requirements Engineering* 1998(3): 182-201 and Ochs M, Pfahl D, Chrobok-Diening G and Nothhelfer-Kolb B (2000) *A COTS Acquisition Process: Definition and Application Experience*, Fraunhofer Institute for Experimental Software Engineering (IESE), ISERN Report 00-02.

[23] This has been pointed out, e.g., by Hannu Harju in a VTT Technical Research Centre of Finland research note from 2002, *Kustannustehokas ohjelmiston luotettavuuden suunnittelu ja arviointi* [Costeffective design and assessment of dependable software], p. 41.

[24] See, e.g., Barry Boehm and Kevin Sullivan, Software Economics: A Roadmap, p. 326. According to the report by the Committee on Information Systems Trustworthiness of the U.S. National Research Council (Schneider,

This trend is natural in comparison to other industries: the market structures supporting software and information system development are rather primitive when compared to those supporting other industries. Even though standard components are increasingly available for software and information system development and every individual part of a system does not have to be developed from the scratch, the designers are still "less able to build systems from specialized, efficiently produced, volume-priced third-party components" as Barry Boehm and Kevin Sullivan note while mapping the future of software economics as part of software engineering discipline in 2000[25]. So, the focus of software business is moving from tailored (where a lot of new programming is essentially needed) to COTS business mode and the component development work is becoming increasingly market-driven instead of being based on contract with one customer. In short, software engineering is maturing as an engineering discipline.

In market-driven software development the component producer concentrates on adapting its products to better reflect the market needs, and not the needs of a particular user organisation in the early stages of software development, as is done in tailored software that is typically developed (analysed, designed and coded) in close

---

*Trust in Cyberspace*, p. 215) the economics of using COTS products and services is irresistible for all consumers, including government, and represents a major shift from the government's historical use of custom-made information technology.

[25] Boehm and Sullivan, Software Economics: A Roadmap, p. 329. The software industry is analogous to other industries in the sense that most software and especially information systems are built from smaller software objects. However, there still is a difference in that the software industry lacks the ability to confidently swap components in and out of systems – the reliability of the replacement component and the systems tolerance towards the new component is difficult to verify as pointed out by Jeffrey Voas in Certifying off-the-shelf software components, p. 53.

cooperation with the customer[26]. The focus of software development has shifted from a customer-oriented view of software functionality towards a supplier-oriented model of saleability[27]. The requirements for software are not tailored for specific needs, but instead they reflect perceptions of a product-marketing organisation about the requirements of a fairly broad market segment[28]. Even the business plans are derived from the anticipated generic demands of a relatively large user base[29].

---

[26] The way software consuming organisations interact with software producers has changed due to the increased product attention. Traditionally, early and close links between users and developers has been considered critical. Today, software consumers and producers use a variety of intermediated means to communicate their needs to developers. For example, packaged software developers build to requirements gleaned from a variety of sources, including help-desk call-log analysis, market research, product reviews, and user groups, of which direct customer contact is one of the least likely means. This argument has been raised, e.g., by Steve Sawyer S in A market-based perspective on information systems development, p. 100. See also Keil and Carmel, Customer-developer links in software development, p. 33–44, already from 1995. This is one of the major dilemmas as Johan Natt och Dag emphasises in his thesis for the degree of Licentiate of Technology in Software Engineering from 2002, *Elicitation and Management of User Requirements in Market-Driven Software Development*, p. 19: the challenges of developing software for larger markets is to satisfy the end user albeit contact with the end user is limited.

[27] This argument has been made, e.g., by Robert Gehring in a position paper presented at the kick-off on a series of influential workshops on economics and information security, *Software development, Intellectual Property Rights, and IT Security*, p. 2.

[28] This concerns also the security requirements of the system as stated by the Committee on Information Systems Trustworthiness of the U.S. National Research Council in its report from 1999 edited by Fred B. Schneider, *Trust in Cyberspace*, p. 110.

[29] This is typical especially in the mature product phase at least in the mass market. In the early phase of the product life-cycle, each customer has to be won on a case-by-case basis, so the needs of the customer are usually considered essential. In the growth phase the concentration shifts more towards identifying customer segments instead of individual customers and the trend is to abandon the idea of serving every customer individually or directly. The argument is made, e.g., Rajala et al. in their review of software business

This, in turn, is highlighting the difference between the development of computer based information systems and market-driven software development. COTS software is a product for the software market. As a business mode it is closer to manufacturing industry than the tailored software business mode that is closer to professional services activity. So, the development, manufacturing, and distribution of COTS software increasingly becomes the work of specialized organisations. At the same time software-consuming organisations increasingly assemble pieces, not build them. Thus, software consumers concentrate on information system development (assembling pieces), while vendors focus on developing packaged products (COTS software).[30] Although the result is that the increased use of packaged components procured from the software product market[31] is changing the focus of software and information system development, the existence of such a market does not mean the disappearance of traditional tailored software or information system development work. Packaged products may not fulfil the needs of many organizations even though there might be possibilities for at least some tailoring (closer to MOTS and parameterised product options) thus making the tailored software business mode more attractive together with in-house development. That is, while the market perspective applies to many forms of development, it is necessary to highlight the market's emergence and resulting changes to the development work, not argue for the market's ubiquity[32]. However, the very existence of such a large and growing market makes

---

models in 2001, Software Business Models, p. 29-30.

[30] Sawyer, A market-based perspective on information systems development, p. 98. Note that it is not just information systems that are developed from software components. Software systems are developed in similar manner and in both of them the development can be characterised of combining either in-house developed reusable software assets or of components bought outside.

[31] Software product market represents the forum for exchanging goods and services between producers and consumers (Sawyer 2001, p. 98).

[32] Sawyer, A market-based perspective on information systems development, p. 102.

(governmental) regulation more likely way of influencing the software and information system development work.

For the sake of clarity, it has to be pointed out that the emphasis in this study is mainly on component developer, unless otherwise stated. What is referred to as market-driven development is thus about the processes of the component developer and not the component user/buyer. However, it has to be made clear that a component user, the buyer of a software product, is typically also a software system or an information system developer, and the development processes of both sides are affected by the focus on COTS products in the software market. There are even considerations of separate third-party component-based software development processes[33].

While the crucial problem from the vendor's perspective is how to make sure that the product meets the demands of the market at the right time, i.e., how to plan product releases, one of the major questions from the purchaser's perspective is how to elicit and verify requirements on third-party software. Market-driven development changes the field also for the purchaser (component user).[34] Much

---

[33] This has been noted in a research project concentrating on software component products of the electronics and telecommunications field and reported by Leena Arhippainen in her graduate thesis for M.Sc. in 2003, *Use and integration of third-party components in software development*. The factors that drive the costs of using COTS software components have been researched, e.g., by the Center for Software Engineering at the University of Southern California (USC/CSE), headed by Barry Boehm. The developed cost-estimating model called COCOTS (COnstructive COTS integration cost model) build on the popular COCOMO II model to predict the effort involved in integrating COTS software products into applications. In the COCOTS model, the assessment activity refers to the process by which COTS components are selected for use. In Reifer, Boehm and Gangadharan, Estimating the Cost of Security for COTS Software, p. 178-186, the research team incorporates the enhancements to the modelling of the impact of security on development effort and duration done in COCOMO II to the COCOTS estimating framework.

[34] Reifer, Boehm and Gangadharan in Estimating the Cost of Security for COTS Software, p. 180, note that current models for predicting the effort involved in integrating COTS software products into applications do not include security as a cost driver. While providing means for estimating the

of what is traditionally viewed as software system or information system development is now opaque to most organizations consuming software[35]. The increasing use of COTS software is causing user organizations to decrease their level of expertise in system development[36]. Since production is separated from consumption, software engineering methods, techniques, and tools are less important to the consumer than is the outcome of their use. That is, vendors are being evaluated by their potential customers on the basis of their products, not their processes. This product focus permeates how vendors develop software and is a fundamental aspect of why packaged (COTS) software development differs from traditional in-house (tailored software) development.[37]

The use of COTS components in software and information system development offer great savings over tailored (custom-written) software: COTS components may be less expensive, have greater functionality especially if the functionality provided is a good match for what is needed, and be better engineered and tested than would be cost-effective for components developed from scratch for a

---

costs of security for COTS software they also (idem. p. 183-184) estimate on the basis of their analyses the percentual increases both for the effort and the duration to the assessment activity (process by which COTS components are selected for use), for tailoring (activities undertaken to prepare the selected COTS packages for use) and for glue code development (development and testing of the connector software, which integrates the COTS components into the larger application).

[35] The development of the mass market for software products has been accompanied by a shift in systems development and expertise from user organizations to vendors as pointed out by Fred B. Schneider, Ph.D. and a professor in the Department of Computer Science at Cornell University, in *Trust in Cyberspace*, p. 188.

[36] Schneider, *Trust in Cyberspace*, p. 198

[37] Sawyer, A market-based perspective on information systems development, p. 101; Carmel and Sawyer, Packaged software development teams: What makes them different, pp. 7-19.

relatively smaller user community[38]. However, it means that developers of an information or a software system (using COTS components) have neither control over nor detailed information about many system components[39]. Software and information system developers are faced with the risks of constructing systems out of unknown black-box components[40].

## 1.3 Security and/or quality – or something in the between?

Security and quality seem easily separable at the first glance. Security in general (applying to the security of people, property, organisations, and even nations, as well as information) means at least two things: a condition in which harm does not arise, despite the occurrence of threatening events; and a set of safeguards designed to achieve that condition. There never can be absolute security in neither of the meanings – every threat cannot be anticipated and the possibility of safeguards failing is always present[41].

---

[38] The testing of COTS components might be better due to the possible extensive field tests in other products and the vendor's ability to hire more and better testers by the help of its size and appeal.

[39] Schneider, *Trust in Cyberspace*, p. 13 and 90; Harju, *Kustannustehokas ohjelmiston luotettavuuden suunnittelu ja arviointi*, p. 51.

[40] This concern has been raised, e.g., by Devanbu and Stubblebine in Software Engineering for Security: a Roadmap, p. 227, and by Jeffrey Voas in Certifying off-the-shelf software components, p. 53. The user of COTS components becomes dependent on a third party (outside vendor) for decisions about a component's evolution and the engineering processes used in its construction (especially concerning quality and security). In addition, the software and information system developer must track new releases of COTS components and may be forced to make periodic changes to her system in response to the new releases.

[41] "The only system which is truly secure is one which is switched off and unplugged, locked in a titanium lined safe, buried in a concrete bunker, and is surrounded by nerve gas and very highly paid armed guards. Even then, I wouldn't stake my life on it." (the original version of this is attributed to Gene

Quality in general has multiple definitions, which all are correct in the specific situations for which they have been developed. One useful definition of quality in relation to information systems and software separates between three distinct components: objective, subjective and non-assessable[42].

*Objective quality* can be said to exist when the product/service fulfils its requirements and other specifications in an objectively measurable way. Process-based quality is a part of this – following good development process will lead to quality[43]. *Subjective quality* is something that the customer expects from the product or service. It is about meeting the customer's expectations of being able to use the product/service for some specific purpose. The subjective quality attribute is needed because it is practically impossible to make an objective evaluation of the level of satisfaction among customers. The third part of quality is concerned with the *non-assessable* components. There are some features in products and services that cannot be evaluated beforehand even subjectively, because they are visible only after some time (e.g. how easily the product can be modified to adapt to changes in not identifiable future customer needs or does the product do what is expected even with unforeseen errors)[44].

---

Spafford)

[42] see, e.g., Wesselius and Ververs, Some elementary questions on software quality control, p. 322.

[43] In software engineering the quality of the product is typically enhanced by improving the software processes. According to Abrahamsson, *The Role of Commitment in Software Process Improvement*, p. 22, software process is seen as a set of activities, methods and practices used in the production and evolution of software. The underlying assumption is that there is a direct correlation between the quality of the process and the quality of the developed software (Fuggetta, Software Process: A Roadmap, p. 27). Software process improvement is concerned with changing the way software development organizations, teams, and individuals perform their work.

[44] Despite the different definitions of quality (which is a theoretical problem), in practice the evaluation of quality is always subjective. Even if the subjective and objective quality could be defined and measured, the evaluation of the results measured is always subjective – the evaluator compares the quality to

Despite the separability of the concepts of quality and security at the surface, the concepts start to blur when going deeper. Vulnerabilities, i.e., defects or weaknesses in the design, implementation, coding, or operation and management of software and information systems that could be exploited to compromise some of the security goals of a system (e.g. confidentiality, integrity, availability) have a lot to do with quality. Poor quality is typical cause of defects and weaknesses. However, these defects become security issues only when they can be exploited, i.e., there is a threat that they might be exploited. In relation to defects in software: a coding bug affecting the performance of the system concerns quality, but a software defect that does not necessarily affect the functionality of the system, but however makes it vulnerable to exploitation, is about security[45].

Information security and quality are thus closely related with vulnerability as the combining factor. With better quality products and services it is possible to diminish the level of vulnerability. Most systems have vulnerabilities of some sort, but this does not mean that the systems are too flawed to use. Not every vulnerability results in an attack, and not every attack succeeds. Success depends on the degree of vulnerability, the strength of attacks, and the effectiveness of any countermeasures in use. If the attacks needed to exploit vulnerabilities are very difficult to carry out, then the vulnerability may be tolerable. If the perceived benefit to an attacker is small, then even an easily exploited vulnerability may be tolerable. However, if the attacks are well understood and easily made, and if the vulnerable system is employed by a wide range of users, then it is likely that there

---

competing products/services and to her own vision of the way the product/service should be or have worked.

[45] This is the reason why it is so hard to test for security. Security flaws do not necessarily appear as a functionality problem – a system can function normally and be completely insecure. These flaws can remain undiscovered until someone looks for them explicitly. This is discussed in more detail in heading 2.3 below.

will be enough benefit for someone to make an attack.[46]

Even though many of the information security problems could be resolved within quality assurance, this alone is not enough, since even a product of extreme quality could be subject of misuse[47]. This is where security functions enter the scene. Security properties (functions or features) such as access rights, security classification and requirements (e.g. confidentiality, integrity and availability) has to be developed into the system. Of course their quality has to be maintained similarly to the development of other system properties, but they improve the security of the system as such (e.g. by defining who is able to access the system, which parts/material and what she can do).

Hereby, a difference between the quality aspects of security (reducing vulnerabilities that could pose a risk to security) and the security properties (features) that can be seen as requirements of software or information systems and that should be made part of the overall product according to user needs is made. The former is a quality issue (also of the security features) and the latter is a

---

[46] This is how Shirey in *Internet Security Glossary*, p. 190-191, (IETF, RFC 2828) clarifies the definition of the term "vulnerability". Information security endangering vulnerabilities can be caused in different phases of the system/software life-cycle: requirements analysis (failures in the sensitivity assessment), development (use of poorly functioning development tools), implementation (not enabling or configuring security features), operation/maintenance (not updating software for security purposes, lack of system audits and monitoring), disposal (improper disposal of information). Vulnerabilities can be located from different components of the information system as Pipkin points out in *Information Security*, p. 68-69: in hardware they can be devastating but contribute only to a small portion of exploited vulnerabilities; a prevalent type and most exploited are found in software; in infrastructure they are typically under the control of outside organisations; in security processes vulnerabilities are typically the target of social engineering.

[47] This means that even if a product is of high 'quality' (meets the expectations of the customer and the specifications) it still might be insecure, e.g. in a situation where the customer did not know to demand security or it has not been considered in the specifications or the component of quality was non-assessable.

functionality issue. The quality aspect of security reduces risks by targeting the avoidance of defects that could evolve into security vulnerabilities. The security features side prevent threats coming from outside, e.g., the attack of computer criminals. The security features prevent the attackers' possibilities to abuse the defects, but the underlying vulnerabilities are remained untouched. The quality aspect of security is more directed towards the problems inside the product while the security features concentrate on threats from the outside, even though they typically are caused by defects in the product.

What is referred to here as 'quality aspect of security' is a gray area somewhere between quality and security. Neither traditional quality enhancement methods, such as basic software development methods and traditional testing, nor the development of security features like access control and other requirements of a software or an information system that should be made part of the system according to user needs alone can solve it. This gray area covers one of the most fundamental problems in information security – security related vulnerabilities in common software packages.

Even though the concentration in this study is on the quality aspect of security, also the development of security features are considered to a certain degree. However, they need to be kept separate in order to understand the problems.

Two other important concepts in security are threat and risk. A *threat* is a potential for violation of security, which exists when there is a circumstance, capability, action, or event that could breach security by exploiting a system vulnerability and cause harm. Threats can be caused by human errors (by users with access and authorization), system failures (due to e.g. of mechanical difficulties in hardware or infrastructure, and unexpected input in software), natural disasters (e.g. fire, flood, earthquake) and malicious acts (attacks and malicious software). The threatening events present in dealing with information security, which are at the same time causes of system and network problems, can be analysed into two categories: accidental and intentional.

Accidental causes of problems in (or threats to) information systems or networks are generally either natural (commonly referred to in the insurance industry as Acts of God or Nature, e.g. fire, flood, lightning strike, tidal wave, earthquake, volcanic eruption) or human but unintentional. The unintentional human errors can be caused by humans who are directly involved (e.g. dropping something, tripping over a power-cord, failing to perform a manual procedure correctly, miscoding information, mis-keying, failing to perform a back-up), by other humans (e.g. unintended cutting of a communications cable during excavation), or by machines and machine-designers (e.g. disk head-crash, electricity failure, software bug). Accidental causes figure prominently in many aspects of trustworthiness beside security, such as safety and reliability[48].

Intentional causes are the result of conscious human choice. Malicious intent is present. Security experts often refer to the efforts of these malicious people as "attacks". Also an attacker – who seeks to cause damage deliberately – may be able to exploit a flaw accidentally introduced into a system. System design and/or implementation that is poor by accident can result in serious security problems that can be deliberately targeted in a penetration attempt by an attacker[49].

Term *risk* is used in many ways, more loosely in social sciences and colloquial language, and with stricter definitions in natural sciences and technology. Common to these definitions is that risk refers to the expectation of loss expressed as the probability that a particular

---

[48] Computer Science and Telecommunications Board, *Cyber-security Today and Tomorrow*, p. 4.

[49] A particularly insidious "accidental" problem arises because of the fact that the precise software configuration on any operational system (including applications, device drivers, and system patches) has almost certainly not been tested for security—there are simply too many possible configurations to test more than a small fraction explicitly. As new applications and device drivers are installed over time, an operational system is more likely to exhibit additional vulnerabilities that an attacker might exploit. Computer Science and Telecommunications Board, *Cyber-security Today and Tomorrow*, p. 4.

threat will exploit a particular vulnerability with a particular harmful result. So, vulnerabilities contribute to risk because they may allow a threat to harm the system. Risks are structurally inside of the processes – they are constructed. In relation with technology they are inherent.

## 1.4 What is regulation?

Because regulation is a widely used and generic concept, there is a need to separate how it is used in this study. The contents of the concept is sought for in order to limit the scope of this study and to make qualifications.

Unfortunately there is as little agreement of what 'regulation' is as there is with 'information security'. Not only is there a conceptual muddle (what term to use), but also a disagreement of the contents of the concepts (what the issue is). However, similar to my earlier treatment of the concept of information security, no formal, universally applicable, definition of the concept of 'regulation' is given. This is mainly because others have already 'mapped' the concept extensively[50]. I can lean on that.

The things that the concept of regulation does are much more important than what it means. In that purpose I will try to show you the issues that the concept of regulation includes and by doing so, hope that I am able to use it to provide some starting points on critical reflections on contemporary problems in information security

---

[50] Doctor Julia Black, a Reader in Law at the Law Department of the London School of Economics (LSE) and the Centre for Analysis of Risk and Regulation (CARR) which is an interdisciplinary research centre at LSE, has made a great effort not only in mapping the concept and its relation to neighbouring concepts such as law and governance, but also in defining the concept of regulation. See, e.g., Black, Constitutionalising Self-Regulation; Black, Decentring Regulation; and Black, Critical Reflections on Regulation. Even though she has specialised on the regulation of financial services, her theoretical insights are more general. Dimity Kingsford Smith provides an appraising evaluation of her conceptualisations in What is Regulation?, p.37-47.

regulation. The understanding of regulation given in here is, following Julia Black, to provide a conceptualisation of regulation that provides the tools of inquiry into the particular problem of vulnerable software[51]. It is used to delimit and construct the scope of the inquiry together with the facilitation of both the analyses and the practical discussions of how regulation affects behaviour and how it might be improved.

At best, regulation is an ambiguous concept. Even the basic regulatory textbooks give at least three definitions[52]. In the first, regulation is the promulgation of rules by government accompanied by mechanisms for monitoring and enforcement, usually assumed to be operating through a public agency (either specially built for the purpose or an existing). This is the way lawyers typically see regulation; simply as a type of legal instrument or as a part of public law. In this sense regulation is a rule of order prescribed by superior or competent authority relating to action of those under its control. In this traditional use of the term it usually denotes a form of intervention that consists of setting and enforcing rules of behaviour for organisations and individuals. It thus contrasts with other forms of state intervention such as public ownership, taxes and subsides or physical alteration of the environment[53]. In practice, discussion of regulation in the narrow sense tends to run into a broader discussion of alternative legal policy instruments, particularly over regulatory reform[54].

---

[51] Black, Critical Reflections on Regulation, p. 19.

[52] See, e.g., Baldwin and Cave, *Understanding Regulation*, p. 2; Baldwin et al., *A Reader on Regulation*, ch. 1. See also Black, Critical Reflections on Regulation, p. 8 and Black, Decentring Regulation, p. 129.

[53] Hood and Scott, Regulating Government in a 'Managerial' Age, p. 1.

[54] See, e.g., Breyer, *Regulation and its Reform* and Ogus, *Regulation*. In considering whether to use law, 'regulation' or some other instrument to achieve a particular policy outcome, 'legalisation' is the consequence of an increasing reliance on law in state intervention, a meaning that must be distinguished from legalisation as general increase of legal norms (as distinct from social custom, convention, or informal social norms) in society as pointed out by

However, there is even a more strict understanding of regulation. According to Black's law dictionary: "Regulations… are issued by various governmental departments to carry out the intent of the law. Agencies issue regulations to guide the activity of those regulated by the agency and of their own employees and to ensure uniform application of the law"[55]. Regulation is rule or order having force of law issued by executive authority or government. Not even the highly detailed parliamentary laws typically used, for example, in Finland, would be included. Only agency regulation and ministerial decrees are considered as regulation under this strict conception.

In the second, it is any form of direct state intervention in the economy or social environment, whatever form that intervention might take. Regulation is any attempt by the government to control the behaviour of citizens, corporations or other parts of the government. In this sense the goal of regulation is often the project of welfare economics: the correction of market failure. In the standard treatments of 'regulation', the 'why regulate' question is nearly always answered in terms of correction of market failures, with the occasional nod to distributional or other ancillary aims[56].

---

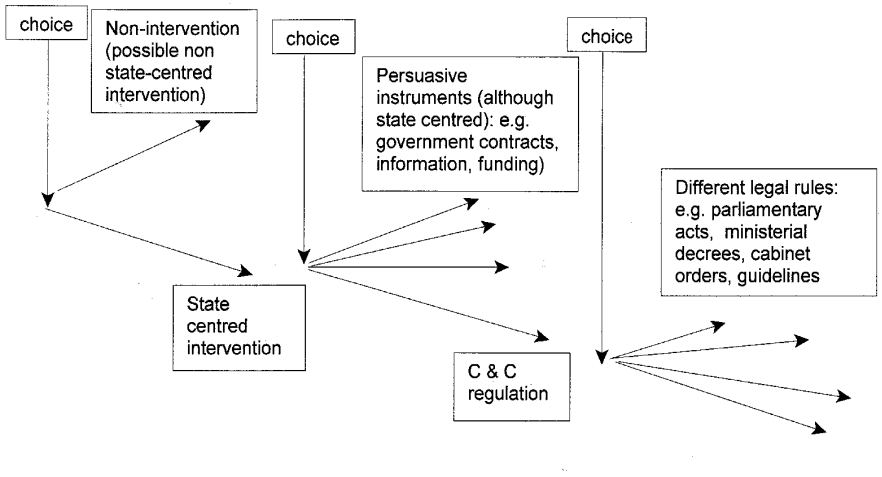
Renate Mayntz in Political Intentions and Legal Measures, p. 57.

[55] Black, *Black's Law Dictionary*, terms 'regulation' and 'regulations'.

[56] See, e.g., Weimer and Vining, *Policy Analysis*, p. 58-159; Breyer, *Regulation and its Reform*; Ogus, *Regulation*, p. 1-121; and Baldwin and Cave, *Understanding Regulation*, p. 9-18. However, as noted by Black in Critical Reflections on Regulation, p. 7, that goal is being displaced, and others added. Notably, the management and distribution of risk: regulating the 'risk society' is a burgeoning academic and policy area and there are signs that existing systems of regulation are coupling the correction of market failure with the management of risk as their organising principle. This is the theme in the analysis of risk regulation regimes in Hood et al., *The Government of Risk*, from 2001. Other goals that regulation ought to pursue, in particular those coming from a socio-legal base, are access to justice (Parker, *Just Lawyers*), or legitimacy (Baldwin and Cave, *Understanding Regulation*, p. 77-85), or the achievement of social justice in some form (Ayres and Braithwaite, *Responsive Regulation*, chapter 3), or the extension of participative forms of policy building into regulation (Black, Proceduralising Regulation Part I and II).

In the third, regulation is all mechanisms of social control or influence affecting all aspects of behaviour from whatever source, whether they are intentional or not. This definition in its scope covers all issues from governmental regulation to everything in social and political sciences. It provides no boundaries where regulation might end and some other influencing factor take effect. Analytical value of this conception is minimal. It is so broad that it contributes nothing.[57]

These definitions can be illustrated with a picture showing the alternatives policymaker has in choosing the tools to be used. The state is just an example of the central authority that performs the regulation. However, it is the most typical one and has the power to use legal measures.




Picture 1-1. The nature of alternatives in the choice of law as a policy instrument. Adjusted from Mayntz, Political Intentions and Legal Measures, p. 58.

[57] Black, Critical Reflections on Regulation, p. 8 and 17. This kind of wide usage of the term 'regulation' in Europe (almost as a synonym for governance) has somewhat hampered the emergence of regulation as a field of study separate from other disciplines. However, it is used in socio-legal studies like that of Lawrence Lessig, The New Chicago School.

In the first choosing point the alternative is between state intervention and non-state-intervention[58]. The possible non state-centred regulations (such as social norms) affect behaviour or there is even a possibility for the state to modify the effects of these regulations. The state can intervene indirectly via these other regulatory instruments and the actors involved. All of this is encompassed by the third definition of 'regulation'.

The second choice concerns whether to use regulatory (public) law typically in the form of 'command and control' regulations or other tools of the state (e.g. taxing, direct funding and other economic incentives, governmental procurement contracting, information, threat of governmental regulation, conscious decision not to regulate). This is what the second definition covers.

The first, and most strict, definition of 'regulation' is visible in the last choosing point; the types of legal measures to use or even in a more limited version, where regulations corresponds to ministerial decrees or decrees given by administrative (regulatory) agencies.

The first two are clearly 'centred' definitions; i.e. regulation is seen to emanate from the state. The usual assumption is that government is the rule-maker, monitor, and enforcer, usually operating through a public agency (ministry, independent regulatory agency, some less independent form etc.)[59]. The second definition keeps to the

---

[58] Note that explicit decision not to intervene, to refrain from regulating, can also be an alternative for legislating as pointed out by Evert Vedung in Policy Instruments, p. 22-23. However, as Jyrki Tala argues in his doctoral dissertation *Lakien vaikutukset*, p. 152, the decision not to intervene has to be explicit and done systematically in order to be seen as a real regulatory choice.

[59] State centrism is the core understanding that many have of 'regulation', i.e., some form of 'command-and-control' (C&C) regulation (regulation by the state through the use of legal rules backed by mainly criminal sanctions). However, C&C has also become a shorthand to denote all that can be bad about regulation as pointed out by Julia Black in Critical Reflections on Regulation, p. 2; the instruments used (laws backed by sanctions) are inappropriate and unsophisticated (instrument failure); government has insufficient knowledge to be able to identify the causes of problems, to design solutions that are appropriate, and to identify non-compliance (information and knowledge failure); implementation of the regulation is inadequate

government as the 'regulator' while broadening the techniques that may be described as 'regulation'[60]. The third definition breaks the connection with the state.

In the conception adopted in this study, regulation is 'decentred', i.e., diffused throughout society. A wider perspective, which deviates from the pure state-centred regulation, is necessary in order to understand the wide area of information security and especially the regulation of secure software development. This regulation is essentially dispersed in different types of self- and governmental regulation, and social norms which have similar and even forceful effects to secure software development. Technologies and methods for their development also play an important role.

Not only is the regulation of information security (even the regulation of secure software development) diffused throughout society, the existing state centred regulation is also scattered in parliamentary laws, governmental regulations (decrees) and guidelines, supra-national regulations such as different sources of EC law (treaties and general principles as primary legislation; regulations, directives and decisions as secondary legislation; general principles of administrative law; international agreements and conventions between member states) and international conventions and guidelines[61].

---

(implementation failure); and that those being regulated are insufficiently inclined to comply, and those doing the regulating are insufficiently motivated to regulate in the public interest (motivation failure and capture theory) including poorly targeted rules, rigidity, ossification, under- or over-enforcement, and unintended consequences. The extent to which C&C does or does not live up to this caricature is an empirical question which has been debated, e.g., by Baldwin in Regulation and by Gunningham and Grabovsky in *Smart Regulation*, p. 38-50.

[60] Black, Decentring Regulation, p. 129.

[61] If it is problematic to recognise what information security is, it is even more difficult to grasp the regulation of it in somewhat a comprehensive manner despite the efforts to coordinate the regulation in different levels. The first attempt for a coordinated policy for on information security in the EU was in 1992 with the Council Decision 92/242/EEC of 31st March 1992 in the field of security of information systems, OJ L 123, 8.5.1992, p. 19-25. More

As a qualification of the subject matter, the notion of decentring is controversial. The recognition that regulation is 'decentred' does not help to limit the scope of issues studied under this label. On the contrary; it extends the concept to cover every form of social control and re-labels almost all questions of social and political science questions as 'regulation'. Moreover, the thing that is doing the regulating is increasingly broadened form the state and some self-regulatory associations to other actors (committees, firms, epistemic communities, contracting individuals) and to other 'factors' such as norms, culture and technology[62]. This has implications on how the 'regulation' is done – what instruments are used, as noted by Julia Black[63]:

> "…if it is *government* that is seen to be the 'regulator' then regulation is used to refer to the use of rules, legal, quasi-legal, non-legal, which may have a certain character (mandatory, facilitatory, performance, technical), which may or may not be accompanied by systematic monitoring and enforcement of sanctions for their breach ('command and control' regulation) by government. Or, … it may refer to any action by government: use of laws, economic instruments, information, persuasion. … *Non-governmental* actors have a similar range of instruments, excluding the legitimate use of force.

---

successful attempts at coordinated policy has been the communications on Creating a Safer Information Society by Improving the Security of Information Infrastructures and Combating Computer-related Crime (COM(2000)890) and on Network and Information Security (COM(2001)298). Despite the difficulty, there have been substantial efforts to give at least a somewhat comprehensive overview of the state centred information security regulation in academic literature especially by Tuomas Pöysti in *ENLIST Information Security Commentary* and by Ahti Saarenpää and Tuomas Pöysti in *Tietoturvallisuus ja laki*.

[62] Black, Decentring Regulation, p. 132-133. With technology, Julia Black, Decentring Regulation, p. 137, is referring to the understanding of and ability to employ, manipulate, or alter the physical or human environment and the products of that understanding. Examples are probability theory (risk analysis), double entry book keeping (audit), design of the built environment and its impact on policing etc.

[63] Black, Decentring Regulation, p. 139.

> Governmental and non-governmental actors may act alone or in any combination. If the *market* is seen as 'regulating' then it is through the interactions of rational buyers and sellers. If it is the broad category of '*social forces*' that is chosen then essentially the analytic tools of sociology are employed: structuring, framing, enabling, co-ordinating, ordering, etc.; if it is '*technologies*' then it is the results of the development and application of understandings of the physical or human environment – the outpourings of the applied, natural, and human sciences." (Italics added)

In an attempt to construct a minimalist core concept of regulation[64] Julia Black has mapped the decentred form of the concept in conventionalist terms, i.e. by looking how the concept is used in practice and ascribing the definition to what the community under consideration (in her case the English-speaking academic and policy community) identifies as 'regulation' (what regulation is used to mean in that particular community)[65]. Such an approach avoids the problems of over- and under-inclusiveness that arise from de-contextualised, generalised abstractions used in essential (identifies central elements of the phenomenon and says that when they are present, then the phenomenon may be termed as regulation – 'regulation is…') and functional (based on the function that regulation performs in society – 'regulation does…') definitions[66].

---

[64] In agreeing that regulation does not solely emanate from state the decentred view (there is similar discussion concerning the 'law' under legal pluralism) leave basically only two possible options for the conception of 'regulation': to abandon any attempt to hold on to a single coherent conception or to attempt to construct a minimalist core concept. William Twining, A Post-Westphalian Conception of Law, p. 206, makes this notion in a similar discussion in relation to the law.

[65] Black, Critical Reflections on Regulation, p. 11-19; Black, Decentring Regulation, p. 133-139. From the breadth of the basis of her argumentation becomes visible that the community she is using in mapping the concept of regulation in conventionalist terms is not limited to English-speaking countries. The academic and policy community she refers to is much wider.

[66] Black, Critical Reflections on Regulation, p. 17-19.

The table where she has included the different uses of the concept perfectly visualises the way 'regulation' is expanding as a concept. Even without further explanations it is useful because it in a compact form shows the different meanings and contents of the concept. Black uses the following five step classification of the different sets of meanings and application of the concept of regulation[67]:

1. what is assumed regulation is (a type of legal instrument, process, an outcome, or a property);
2. who or what is performing it (state institutions, non-state institutions or actors, economic forces, social forces, or 'technologies');
3. what institutional or organizational form the regulation is assumed to take (e.g. ministries, supra- or international bodes, associations, firms, networks, market, norms, language);
4. with respect to what actors or areas of social life is it occurring (firms, markets, family, health, education etc.) and
5. how regulation is conducted, through what mechanisms, instruments and techniques (e.g. rules, taxes, trust, interaction of rational actors).

---

[67] Black, Critical Reflections on Regulation, p. 12; Black, Decentring Regulation, p. 134-135.

| **Table 1: Regulation – an ever expanding concept** | | | | |
|---|---|---|---|---|
| **(A) What is regulation?** | **(B) Who or what does it?** | **(C) What form does it take?** | **(D) With respect to what actors or area of life?** | **(E) How is it done, via what instruments/ techniques?** |
| *Type of legal instrument*<br><br>*Process of*:<br>• 'controlling, governing, or directing' (OED)<br>• 'altering or controlling with reference to some standard or purpose' (OED<br>• enabling/facilitating<br>• co-ordinating<br>• influencing<br>• conferring a pattern on something, ordering<br>• rendering constant<br><br>and the *process* is:<br>• intentional<br>• goal-directed, problem-solving<br><br>An *outcome* – the result of the interaction of actors/networks/'forces'<br><br>A p*roperty* of self-correction<br><br>A p*roperty* whereby the nature and growth of parts of an organism are interrelated so as to produce and integrated whole enabling adaption (biology) | State institutions (regional, national, 'extra-national) | - ministries, departments, agencies<br>- supra-national bodies (EU)<br>- international bodies (WTO)<br>- courts | - economics (firms, markets)<br>- any other (family, education, health, government, etc.) | - rules (legal, quasi-legal, non-legal, universal, sectoral, bilateral)<br>- other instruments (financial, market based, information)<br>- monitoring<br>- sanctioning |
| | Non-state institutions/ actors | E.g.<br>- associations<br>- committees<br>- firms<br>- individuals<br>- epistemic communities<br>- networks | - economic<br><br>- any other | - rules (legal, 'quasi-legal', non-legal; multi-lateral, bilateral, unilateral)<br>- other instruments (financial, market-based, information)<br>- monitoring<br>- sanctioning<br>- trust |
| | Economic forces | Market | - economic<br>- any other | - interaction of rational actors |
| | 'Social forces' | E.g.<br>- norms<br>- institutions<br>- language<br>- cognitive frames<br>- culture<br>- systems<br>- networks | - economic<br><br>- any other | E.g.<br>- structuring<br>- framing<br>- enabling<br>- co-ordinating<br>- ordering<br>- translating<br>- self-referential reproduction |
| | 'Technologies' | Understanding of and ability to manipulate physical and human environment | - any | - products of those understandings, e.g. statistics, probabilities, engineering, IT |

Table 1-1. Regulation – an ever expanding concept. Presented by Julia Black in Critical Reflections on Regulation, p. 12, and in Decentring Regulation, p. 134-135.

As Julia Black has noted, the way 'regulation' is conceptualised depends heavily on the problem or issue that the writer is focussing

on[68]. The understanding of what regulation is dependent on what we want to do with it. If it is to serve as a descriptive device for an empirical investigation into what structures or constrains the behaviour of individuals, organizations, or systems, then a wide-ranging conception of regulation is needed[69]. Such a definition would probably come close to embrace everything on the table above as being part of the concept of regulation. This turns almost all questions of social and political science into questions of 'regulation'.

The purpose of this study being the examination of the possibilities of 'regulation' to guide behaviour, and not just the description all the factors and actors that influence behaviour, leads to the concentration on the intentional attempts to control or order behaviour of others. The whole notion of regulation guiding behaviour implies purposive

---

[68] Black, Critical Reflections on Regulation, p. 9 and 19; Black, Decentring Regulation, p. 141-3.

[69] This is how Professor of law at Stanford Law School and the founder of the school's Center for Internet and Society Lawrence Lessig, one of the most influential person behind the change of the legal culture of IT professionals towards the possibilities of regulation in cyberspace, uses the concept in making his argument for regulability of 'code'; eventually that 'code is law' (Lessig, *Code and Other Laws of Cyberspace*, Chapter 7 and Appendix; Lessig, The New Chicago School, p. 661-691). This different usage of concepts in Lessig's work makes it really difficult for a lawyer to see the differences in law, regulation and governance. What Lessig is mainly talking about when he says that 'code is law' is some form of governance or regulation in the widest sense, not law in any strict sense. However, his aim is not conceptual preciseness but efficient rhetorics, and in that he succeeds with the analogy between code and law. However, this type of analogy makes it really difficult to discuss and critice his theses, as the notion made by Syme and Camp, *Code as Governance, The Governance of Code*, p. 24, of the flawfullness of the analogy code and law shows. They consider the analogy flawed with the notion that a person has freedom to choose to bind herlself to a certain code, but in relation to law there is no such freedom. What is unseen is that in many other forms of governance or regulation (other than law) this kind of freedom exists. And this freedom exists even in relation to law, but in a more restricted sense and concerns only those with the ability to move from a jurisdiction to another (e.g., global enterpises).

action as explained in the philosophy of law[70]. This is why I adopt the following definition of regulation presented by Julia Black[71]:

> *Regulation is a process involving sustained and focused attempt to alter the behaviour of others according to defined standards and purposes with the intention of producing a broadly identified outcome or outcomes.*

This definition disconnect regulation from the sole activity of government and enables the decentred understanding; also others than the governmental actors can and do have an intention to attempt to control the behaviour of others. Regulation is also a purposive activity; its orientation is on problem-solving or goal-attaining in the sense there are outcomes the regulation as a process (activity) is intended to produce[72]. Having intentionality as a defining feature excludes certain parts of the table set out by Black; the categories of market forces, social forces and technologies. There are no 'actors' having intentions to constrain behaviour or, as in the case of market forces, the outcome of the interactions of actors is not intentional. The temporal dimension (sustained and focused attempt) limits the extension of the activity in time.[73]

> The cybernetic understanding of the mechanisms of regulation (standard setting, information gathering and behaviour modification)

---

[70] See, e.g., Joseph Raz, On the Functions of Law, p. 284; Christina Redondo, *Reasons for Action and the Law*, p. 14.

[71] Black, Critical Reflections on Regulation, p. 20. For a deeper discussion on the definition see also Smith, What is Regulation?, p. 37-47.

[72] As Black notes in Critical Reflections on Regulation, p. 21, this does not mean, however, that regulation in general is the *sole* the project of, for example, welfare economics, or addressing the concerns of the risk society. The value base and the justifications for regulation have to be considered more extensively.

[73] Black, Critical Reflections on Regulation, p. 19-20; Smith, What is Regulation?, p. 42-43; Parker et al., Introduction, p. 1-2.

> is left out from this definition[74]. I do share Black with her fear of regulationists interpreting regulation too narrowly, e.g., only as standard setting, and concentrating only on the executive or legislative process of setting standards. However, due to behaviour modification being the goal of regulation and the research interest in this study on being on the effects this process assumes to cause on its targets, the other two mechanisms (information gathering and standard setting) are combined into the behaviour modification function. They actually are a part of the regulatory toolkit (ways of modifying behaviour) as we will see when studying the ways regulation affects behaviour[75]. The way standards are set and communicated, and the methods that are used in information gathering affect behaviour. There actually are separate types of regulatory tools that employ the capacity of the information resources of the regulator and that are based on the ability to authoritatively and legitimately set standards, goals and objectives.

This definition, so I believe and hope that you agree as we go along, helps to see the tools of inquiry to examine intentional attempts to alter the behaviour of those developing software and information systems. The assumption of at least some intentionality involved in the conception of regulation used – the intention to direct behaviour, even if there might be unintended consequences or side effects – makes it possible to distinguish regulation from all other questions of social control and ordering.

## 1.5 Two combined perspectives into the study of regulation

The need to be aware of the perspective when studying the ways regulation influences behaviour has been emphasised in regulatory

---

[74] More about these mechanisms and the cybernetic understanding of regulation see e.g., Black, Enrolling Actors in Regulatory Systems, p. 67-69 and Hood et al., *The Government of Risk*, p. 23-27.

[75] See section 3.1. This is also noted by Black in Enrolling Actors in Regulatory Systems, p. 69.

studies[76]. Every intentional attempt to affect behaviour includes at least two perspective; that of the decision-makers (regulators) or that of the objects of the regulation[77]. There always is the person making the decision to affect the behaviour of the target group and the person experiencing the constraint of the regulation regardless of the form of the regulation[78]. The choice of the perspective affects several issues.

*First.* The role of intentionality requirement in regulation is dependent on which one of these perspectives is taken[79]. When taking the more usual point of view of the regulator the question is what tools are available to it to solve a particular problem. The questions of the most effective and efficient means to achieve the desired objectives of the regulator, the costs of different regulatory alternatives, the restrictions and limits in their use become important. Regulation is an activity (with actors) and the intention of that activity (and actors) becomes an issue.

From the point of view the regulated (the object of regulation) the question turns into what forces are they subject to[80]. Citizens,

---

[76] Black, Critical Reflections on Regulation, p. 10; Tala, *Lakien vaikutukset*, p. 37 and p. 384-385 in the English summary.

[77] Note that implementation research also emphasises the perspective of the field level implementation agents (Mazmanian and Sabatier, *Implementation and Public Policy*, p. 11-13).

[78] In legal education the concentration is typically on the perspective of the regulators, or more precisely, on the perspective of one of the regulators, that is, the judge. Lawyers are taught to give interpretations of the law that are valid. Their analysis starts from the legal system, from the normative world, and its interpretation. The user perspective into the law, as Håkan Hydén calls it in *Rättsregler*, p. 22-23, on the other hand starts from the real world and concentrates on analysing what the valid law means for a certain economic or social action. The focus is on the consequences that the law is expected to have on the practised area.

[79] Black, Critical Reflections on Regulation, p. 10.

[80] In addition to the intended target group (if there is such) the concept of 'objects of regulation' covers also non-intended ones that might be subject to unintended effects. This distinction is made, for example, by Vilhelm Aubert in his study of the social functions of law, *Rettens sosiale funksjon*, p 124. In the final end it is the individuals as citizens or consumers, part of a societal or a

associations, corporations, courts, administrative authorities etc. as objects of regulation are more interested in what the regulations give, demand or constraint, what possibilities it gives and restrictions it produces than in the tools used. From this perspective regulation is an *outcome* and the intentionality element is less relevant; the constraint is experienced regardless whether the 'regulator' (any of the 'doers' in Blacks table) intended to direct behaviour or not.

*Second*. The perspective affects the conception of the regulators – parties able to affect the behaviour being regulated. The objective point of view easily leads to elide the other actors and factors in the problem field. Concentration on the specific decision-maker distracts attention from the relationships between different actors and their possibilities to influence. The point of view of the objects of regulation is better in recognizing the actual operational environment where one actor and its instruments are just that – one factor worth noticing.[81]

*Third*. The choice of the perspective also affects the way (the processes under which and mechanisms by which) the effects of a specific regulation are understood to come into being (to be generated in technical terms). Here the decision-makers point of view often causes an implicit acceptance of the hierarchical, top-down commanding, model of the mechanisms by which regulation affects

---

commercial group, or as an official (part of a public authority) that make the decisions. However, as part of a larger group the individual is subject to the rules and norms of these institutions, thus attenuating individual preferences or desires.

[81] Tala, *Lakien vaikutukset*, p. 40-41; Black, Critical Reflections on Regulation, p. 10.

behaviour[82]. The subjective point of view makes it easier to see that this is an oversimplified understanding[83].

*Fourth*. As Tala differentiates between the four factors that shape the effects of a law reform and that in relation to which the effects of legislation should be examined (objectives, substance, implementation, and reaction of the objects of a law reform) he also makes clear that in their use specific perspectives are stressed[84]. Analysis of the objectives and substance of regulation emphasises the perspective of the decision-maker and the study of the implementation and the reaction of the objects of regulation stresses the point of view of the objects.

In this study, the point of view of the regulator and the regulated are used in the following way[85]. My concentration is somewhat emphasised on the decision-makers even though the point of view of the objects of regulation is also taken; regulation is seen both as an activity, and (but not merely) as an outcome. The former perspective is stressed due to the emphasis being on the substance of regulation as a base of analysis of the regulatory capacities of regulation. The decision-makers perspective is also emphasised in the formation of the concept of regulation on the basis of the intentions of the regulator.

---

[82] Tala, *Lakien vaikutukset*, p. 40. This hierarchical model, according to which the objects of regulation follow the orders of the regulator backed by sanctions in a uniform and predictable way, is a portrait of the way the legal order is seen to cause effects. This model is behind much of command-and-control regulation. (Tala, *Lakien vaikutukset*, p. 22)

[83] The unintended effects of regulation are easier to make explicit with the perspective of the regulated because the spectrum of the regulators is wider and the understanding of the mechanisms by which the effects are generated is more diverse.

[84] Tala, *Lakien vaikutukset*, p. 41-42 and 351-352 in Finnish, and p. 384-386 in the English summary.

[85] As Jyrki Tala in *Lakien vaikutukset*, p. 41 (p. 384 in the English summary) points out, these perspectives should be used together when studying the effects of legislation.

By also considering the role implementation and the assumed reactions of the objects in considering the capabilities of regulation to affect behaviour, the perspective of the objects (the regulated) is taken along. The perspective of the regulated is stressed especially in the analysis of the ways regulation causes effects on behaviour and basing the mechanisms on the conception that the effects of regulation become existent only through the decisions (either conscious or not) of individuals. In short, the perspective is that of an outside researcher that balances between the perspective of the regulator and the regulated.

## 1.6 Research questions, purpose and contribution

This research focuses on analysing the capabilities of regulation to influence secure software development, i.e., the influence mechanism of instruments and the factors that shape their influence. In that purpose, the reasons for lack of security in COTS are analysed and feasible regulatory tools that could solve the problems are sought. The goal is to get an understanding of the possibilities of the most attractive regulatory instruments to influence secure software development and to develop an approach that would enable the evaluation of other instruments.

Assuming the importance of the security of the packaged software components (software products) for the security of our current form of society, this study concentrates on analysing alternative regulatory means, i.e., regulatory instruments, designed to achieve the desired state of affairs. The primary interest is in evaluating alternative regulatory means to achieve a state of affairs where more secure software would be available in the market. In short, the research question of this study is: *Can the effects created by the most plausible regulatory instruments contribute to secure software development and, especially, do they correlate with the social objective of achieving secure software?*

This is done in order to benefit regulatory drafting, i.e. the preparation of different regulatory instruments, by making explicit

the different types of instruments that can be used in the regulation, what their possibilities are in enhancing secure software development and how they are likely to affect[86]. In doing so, I come close to the type of research that is needed for planning and decision-making in social matters[87].

When analysing the regulatory capabilities of different instruments the purpose is to identify problematic issues in regulation that need correction. The purpose is also to provide partial information for the evaluation of which instruments work best. Information on the capabilities of different regulatory tools to influence secure software development is a part of the material needed for the comparison of the desirability of different regulatory instruments. An analysis of the incentive mechanisms and possible effects provide information about the possibilities of different regulatory instruments to enhance the

---

[86] Note the difference to traditional dogmatic legal research where the information produced is considered to assist the judge in her decision-making about valid law and the arguments used in the decision-making. In the regulatory approach used in this study the purpose is to provide information to the use of regulators by considering the advantages and disadvantages of several regulatory alternatives. No single solution is searched for.
The need for a new type of legal scholarship serving not only lawyers acting in the role of judges, but also as advisers or legislators, has been raised in an emerging discipline that applies the tools of legal theory to legislative problems. See, e.g., Luc Wintgens, Legisprudence as a New Theory of Legislation. In the German tradition this has been emphasised already in the 1970's, e.g., by Peter Noll as pointed out by Jan Hellner in *Lagstiftning inom förmögenhetsrätten*, p. 147 (with reference to Noll P (1973) *Gesetzgebungslehre*, Reinbek bei Hamburg, Rowohlt Taschenbuch Verlag, chapter 1). Unfortunately my lack of German language skills have prevented me from going deeper into the the extensive German language literature following the work of Peter Noll and Jürgen Rödig that has aimed at providing principled tools for practical law drafting. For references and a short overview, see Hellner in *Lagstiftning inom förmögenhetsrätten*, p. 145-155.

[87] This is what Aulis Aarnio in Some Conceptual Foundations of Legal Policy Research, p. 225-230, refers to as legal policy research. The main research interest in this study is on the third type of information the policy research should provide according to Aarnio (idem., p. 229-230), i.e., information concerning the alternative means leading to the desired final state of affairs.

development of secure software. This information is useful for the regulators in recognising the issues to be enhanced and, at minimum, not to be hindered when pondering the need for regulation. It is needed for the decision-making of the desirability of regulatory instruments, and helps regulatory drafting by making explicit the possibilities and limitations of the instruments.

The difference to the client-oriented advice giving relevant to public decisions, i.e., the professional activity called policy analysis[88], is that the purpose is not to give direct advice in public policy, but to analyse different means to enhance certain public policy, i.e., that of secure software development. Full evaluation of the feasibility and desirability of different regulatory solutions for the problems of developing secure software would require information much more widely than just of the regulatory capabilities of instruments. Issues like the efficiency, equity, manageability, legitimacy and political feasibility of the specific instruments or their mixes would be relevant.

However, such an analysis is not feasible in an academic study. It implies the need of contextual normative and political arguments which are the work of the regulator and not the researcher. The analysis of the regulatory capability of specific instruments is not a sufficient basis for regulatory decision-making[89]. At best, it is only a guide to improve the quality of political and administrative decision-making.

Note that this is neither a pure prediction of the effects of regulation nor an empirical verification of effects conducted after the enactment of a specific regulatory instrument. Even in the evaluation of the effects of regulation on secure software development based on existing research, the concentration is primarily on the type behaviour the regulation authorises or proscribes rather than on

---

[88] For a definition see the basic textbook of David L. Weimer and Aidan R. Vining, *Policy Analysis*, p. 27-28.

[89] Similar role has been assigned for regulatory impact analysis in general in a collection of best practices in OECD countries (OECD, *Regulatory Impact Analysis*, p. 7).

attempting to make estimations of the probability of behavioural changes, i.e., what behaviour the rule will induce or discourage[90]. The analysis concentrates more on the types of incentives for behaviour provided by the regulation, than on the probability of actual effects. The analysis focuses more on the incentives regulation provides for secure software development than on estimating the probability of actual behavioural changes.

Similar to the differentiation between intended and unintended consequences, sociologist Guy Rocher has drawn an illustrative distinction between law's "efficacité" and law's "effectivité"; that is, law's capacity to produce intended results and law's capacity to produce social consequences in general[91]. In this study the concentration is on the ability of regulation to influence secure software development, its ability induce, under the given circumstances, such behaviour as will bring about the desired change in software development towards more security consciousness, i.e., the efficacy of regulation in enhancing secure software development.[92]

This study contributes by adding a regulatory approach to the recently emerged international discussion on the economic aspects

---

[90] This is the way MacCormick in On Legal Decisions and their Consequences, p. 239 and 254, understands to the role of consequentialist reasoning in judicial decision-making. I.e., judicial decision-making is not, and according to MacCormick should not focus so much on estimating the probability of behavioural changes, as on possible conduct and its certain normative status in the light of the law.

[91] Rocher, L'effectivité du droit, p. 133.

[92] However, since secure software development is not an express objective of regulation, instead it is the central argument of this study that it ought to be, unintended consequences cannot be bypassed. This is why both intended and unintended effects, anticipated and non-anticipated, and those that appear in the subject field of the regulatory instruments or outside of it are still of interest. According to Tala (*Lakien vaikutukset*, p. 66-70) and Ervasti and Tala (*Lainvalmistelu ja vaikutusten ennakointi*, p. 12) this is the approach used in the analysis of the effects of laws. Since the analysis concentrates on the capability of regulation to influence an abstract societal goal (secure software development), of the variety of effects regulation can produce only those that have relevance for secure software development are gathered.

of information security. Not just a deeper understanding of the role of economic considerations in secure software development is gained, but also a map of possible regulatory solutions and their capabilities in influencing secure software development is achieved.

In order to be able the find answers to this general research question following supporting research questions have to be answered first[93]:

First, *what are the current incentives for secure software development in highly competitive software product markets?* In order to be able to analyse the capabilities of regulation to influence secure software development, a rough understanding of the incentives for secure software development has to be gained. However, the making of a somewhat comprehensive picture of these incentives is only a secondary purpose of this analysis[94]. The primary purpose is to find the correct assumptions about the factors causing problems and about what must be changed in order to solve it. This is information is needed to find the regulatory solutions that can solve the problems, i.e., causally adequate instruments[95]. Reasons and explanations for the widely

---

[93] These supporting research questions are based on the idea made expressive by Mayntz in The Conditions of Effective Public Policy, p. 127, that underlying the design of efficacious policy program there must be an adequate causal theory. There must be correct assumptions about the factors causing a problem and about what must be changed in order to solve it, and how regulation can be used to effect behaviour changes. Note that Mayntz uses the term effectiveness instead of efficacy in talking about the ability of a policy program to induce behaviour that will bring about the desired change.

[94] This research question tries to yield information about the original conditions prevailing the policy decision on secure software development. The first type of information that [legal] policy research should provide according to Aulis Aarnio, Some Conceptual Foundations of Legal Policy Research, p. 229-230.

[95] As Mayntz points out in The Conditions of Effective Public Policy, p. 127, causal adequacy can only be defined in relation to a given problem situation. It is not a formal characteristic that can be recognised by looking at the types of instruments used in themselves or the whole policy program itself. This makes it difficult, according to Mayntz potentially even impossible (idem.), to utilise the most interesting and important aspects that policy analysis and especially implementation analysis have provided from several fields of (especially public) policy, i.e., adequacy of the underlying causal theory and other theoretical

acknowledged insecurity of the packaged software used extensively are sought. Preliminary suggestions of possible solutions to the identified problems in the literature are also sought.[96]

Second, *how does regulation guide behaviour or how can it influence behaviour?* Without an understanding of the ways regulation guides behaviour the regulatory capabilities of different instruments is difficult to see and impossible to analyse critically[97]. This is a somewhat neglected area, for the reason why a little deeper understanding of the ways regulation influences behaviour regulation needs to be provided in order to find the necessary heuristic tools for analysis. Of interest are the mechanism by which regulation affects behaviour, who is doing the regulation and the actors to whom the regulation is directed.

This part of the study contributes to regulatory theory and especially the analysis of effects of regulation. Insights form various disciplines make explicit several heuristic tools for the analysis of the effects of regulation. During the study, the applicability of these heuristic tools for a wider regulatory perspective and context is tested and insights into their usefulness made.

---

assumptions underlying a policy program. The adequacy of the theoretical assumptions underlying a program has been identified as some of the most interesting and important aspects of implementation analysis by Mazmanian and Sabatier in *Implementation and Public Policy*, p. 11.

[96] The deeper and more coherent understanding of the problems in secure software development gathered forms the basis for the identification of possible instruments. This information is not new to those participating in secure software development, but as a basis of regulatory analysis this information has been too scattered and just recently emerged into the serious discussions of changing practices of secure software development.

[97] The interest in the influence mechanism is purely instrumental. The mechanisms by which regulation affects behaviour are refined only in order to be able to analyse the capabilities of regulatory instruments.

## 1.7 Of method and material

The method of literature review was used in the analysis of the first supporting research question in chapter 2 where the concentration is on the current incentives for secure software development and the incentive structures of the actors. In this chapter an understanding of the reasons for insecure software is provided; the main problems are seen to be stemming from the network economic environment where software product development is done. The literature reviewed includes security relevant material on software economics and engineering, information systems science, computer science and textbooks from information security practitioners. The scattered analysis of the disincentives for secure software product development is gathered and linked together with the economic arguments of the motives for secure software development. In addition to library catalogues, the material was sought from journals, on-line publications, conference proceedings, web pages of relevant institutions, online course-material and personal web pages of the authors.[98]

Note that since my analysis relies on literary sources and concentrates especially on academic research, there necessarily is a time lag in relation to the present methods of secure software development. In this sense, the analysis stems from an abstract situation where the quality aspect of security has not been of much interest and no major effort has been put into improving the security of software. The argument is that the increasing interest in secure software development during the last couple of years together with the consequent improvements is partly due to the regulatory pressure; i.e., software vendors desire to improve the security of software has partly been due to the increased threat of further regulatory action.

---

[98] The material reviewed is visible in the references of the heading 2.

A change in the method is required when turning to the ways regulation affects behaviour in chapter 3[99]. The method used in answering the second supporting research question is that of analytic isolation. When studying the reasons for actions (incentives) provided by a regulation itself it is not possible to proceed from the wide angle of individual decision-making and its comprehensive depiction; also other factors than just the regulatory instrument in question affect the behaviour and choices of individuals. The regulation is just a part of a wider entirety of influencing factors formed by different individual or social issues that direct behaviour (e.g. motives, beliefs, aims, hopes, emotions, knowledge)[100]. The regulations not necessarily are even the explicit cause of influence as noted by Tala[101]. The incentives provided by the regulatory instruments themselves need to be isolated. Otherwise the incentives get lost into the general motivations for behaviour, and the reasons for action provided by the regulatory instrument as such are lost.[102]

---

[99] In chapter 3 heuristic tools are provided for the analysis of the influence mechanisms. Also classifications both of the instruments used in regulation and of the possible regulators are made on the basis of the wide understanding of regulation not just being state-centred. Considerations of the role of law in the process of regulation are also included.

[100] This is typically acknowledged in sociolegal studies, like that of Mathiesen, *Rätten i samhället*, p. 30-31, and in legislative theory (see, e.g., Tala, *Lakien vaikutukset*, p. 17-21). As MacCormick notes in his philosophical inquiry into the proper role of consequentialist reasoning in juridical decision-making (MacGormick, On Legal Decisions and their Consequences, p. 254), legal rules and rulings in law are not *causes* of behaviour – they are only grounds for choice by people. This implies that the law and rulings in law are not the sole grounds for decision-making.

[101] Tala, *Lakien vaikutukset* 2001, p. 287.

[102] As Frederick Schauer, *Playing by the Rules*, p. vii, in his philosophical study on rule-based decision-making that draws from an extensive literature base in philosophy and legal theory correctly points out, analytic isolation as a method is superior to that of trying to capture the full complexity of behaviour. Despite such aspirations for comprehensive accuracy giving faithful depictions of reality, their accuracy often fails to increase our understanding since they tend only to replicate the vagueness and the messiness of life.

The way to isolate and identify the reasons for action provided by regulation itself is the exclusion of human behaviour and other motivational factors than those provided by regulation as far as possible; to analytically isolate the influencing mechanisms of regulation from all the other influencing factors[103]. The method used in answering the second supporting research question is isolation of the mechanisms by which regulation influences behaviour and the reasons for action it provides of all the possible influencing factors for human behaviour. The analysis of the ways regulation affects behaviour is done at a theoretical level and leans heavily on prior analysis done in legal theory and the philosophy of law, sociology (especially law and sociology) and political science.

It is illustrative to distinguish between the classifications of normative and social functions of norms[104]. It is a separation between the ways regulation affects behaviour and the effects it either has or intends to have. The normative functions are the ways in which regulation guides human behaviour. It is common to all norms, legal or otherwise, that they guide human behaviour; they are reasons for performing or abstaining from a certain action. The normative

---

[103] The sociology of law has largely foregone the experimental method of isolating the effects of law from all the other influencing factors especially when studying the intended effects; as the Professor of Sociology of Law at the University of Oslo Thomas Mathiesen points out in *Rätten i samhället*, p. 59, the effects of laws cannot be separated from the multitude of other influencing factors. The main conclusion of the sociology of law in relation to the intended effects of laws is that when the law is placed in a social, economic and political context that supports its intentions, then it is going to effect as intended (idem. p. 60). This does not, however, render the isolation of the influencing mechanisms as futile or impossible. It only means that in the evaluation of the effects of regulation such an approach is not feasible. When analysing regulatory capacity, the reasons for actions regulations provide, such an approach is essential.

[104] The distinction between the normative and social functions has been made especially in relation to law by Joseph Raz in On the Functions of Law, p. 278-305. The distinction between normative and social functions is not meant as a classification of functions. It is rather a distinction between types of classifications, between different principles of classification of functions.

functions are the reasons for human behaviour provided by the regulation. In considering the variety of these reasons the actual behaviour of people is immaterial; the question of why people comply with regulation and the other influencing factors are closed as far as possible out even though their existence is recognised while analysing the normative function.[105]

Social functions, on the other hand, are attributed to regulations because of the social effects they have or are intended to have. The social functions are determined by the social effects regulation causes or can cause. Regulations fulfil their social functions because of their particular normative character; the normative function is part of the means by which regulation performs its social functions. The social functions, the effects created, depend on the compliance of the addressees, the application of the norms and the effects of the existence of the norms on human behaviour (attitudes, etc.).[106]

In the analysis of the ways regulation influences behaviour the classification principle of normative functions is used. The question is answered by trying to see how normative guidance communicates itself into the behaviour of the objects of regulation, or, in a rephrased form, what is the structure of the mechanism used to direct behaviour[107]. The normative analysis, by adopting the perspective of the regulator, supposes that a regulatory instrument (especially law)

---

[105] The other factors that provide reasons for action and the behaviour of the norm addressees is relevant only at the level of the basic theoretical assumption that norms are generally followed and enforced (i.e. the regulatory system is generally efficacious). To the degree this necessary, but not a sufficient condition for the existence of a legal system (if it were largely ineffective it would fail to provide reasons for actions and would not guide human behaviour) is correct will determine the strength of the reasons provided by a law. The other factors that influence behaviour beside law are further disregarded. (Raz, On the Functions of Law, p. 280, 282 and 287)

[106] Raz, On the Functions of Law, p. 280 and 287.

[107] Kauko Wikström makes a similar qualification in his study about guiding with legal norms, Ohjaaminen oikeusnormeilla, p. 401. He concentrates purely on the normative guidance provided by legal norms.

is a reason for action because of its content[108]. Thus, the concentration is on the substance of the norms, the guidance they provide for behaviour and the mechanisms used in this. This normative analysis is done in section 3.

However, this normative analysis forms only a part of the mechanisms that shape the effects a regulation can have or can be predicted to have[109]. The normative guidance provided by regulation influences the effects regulation can have, but is not the only factor. The pure normative and objective analysis makes explicit only the means by which regulatory instrument are assumed to influence behaviour. It does not say much about the factors that shape the effects of a regulatory instrument in practice[110]. A pure normative remark is made, not a psychological one[111].

---

[108] Raz, On the Functions of Law, p. 282. Redondo in *Reasons for Action and the Law*, p. 80-81, provides a critical evaluation of this approach.

[109] In legal scholarship, it is often incorrectly assumed that the content is also a sufficient indication of the effects produced by the norm as Jyrki Tala correctly points out in *Lakien vaikutukset*, p. 35.

[110] The analysis is made from the perspective of the regulator; it is the regulator that assumes how the regulatory instrument affects behaviour. The different factors that influence behaviour beside the regulation are excluded. In the second research question the perspective of the regulated are taken along.

[111] Raz, On the Functions of Law, p. 284; Redondo, *Reasons for Action and the Law*, p. 41. As Moore, Authority, Law, and Razian Reasons, p. 842-843, notes in discussing the central presupposition of Raz's theory of authority and law of the kinds of reasons for action that authoritative legal norms give and especially in discussing the types of reasons given by morality from the perspective of the philosophy of the law, the objective reasons which morality gives to all rational agents need not be subjective reasons. Raz is not charting the structure of our actual psychology when we reason practically. Instead, his charting of the structure – the systematic relations that hold – of this reason-giving nature of law and morality helps us to analytically separate the specific ways in which regulation influence behaviour apart from all the other influencing factors. By concentration on the normative functions the political is separated from the legal or regulatory. The concentration is on the inner logic of the influence mechanism. However, as Moore, Authority, Law, and Razian Reasons, p. 843-845, makes explicit, the structure that Raz gives to objective moral reasons parallels exactly with a familiar picture of the

A richer overall picture is needed in order to understand the influencing capability of a specific regulatory instrument and to answer the main research question. The analysis of the efficacy of the two different regulatory instruments selected for deeper substantive study of a large variety of instruments from different regulatory strategies – vulnerability disclosure in chapter 4 and software product liability in chapter 5 – necessarily also takes along the behaviour of the norm addressees (the objects of regulation)[112]. The effects created depend on the degree to which laws are obeyed and applied and on the effects of the existence of laws on human behaviour, attitudes, etc.[113]. The method of analysis for the primary research question uses the classificatory principle of social functions and necessarily takes along the political aspect.

The normative analyses of the influence mechanisms of the regulatory instruments utilising the theoretical considerations done in chapter 3, where the intention is to make explicit the mechanisms by which the instrument affects secure software development, thus forms only a basis for the analysis of the efficacy of the regulatory instruments. In the consideration of the consequences attached to actions, i.e., additional reasons for action provided by the regulation, the focus is on those consequences that are especially created by or employed with the help of the regulatory instrument be they legal or non-legal. Consequences that are employed otherwise are not included; instead, they are analysed in the second part of the chapters 4 and 5 where the factors that shape the influence provided by an instrument are analysed.

Because the pure normative approach is not sufficient to make visible the efficacy of regulatory instruments, the analysis needs to

---

subjective reasons on which persons act.

[112] These two were chosen for a deeper analysis due to their attractiveness in tentative analysis. Those practising information security had raised these two, among others, as feasible alternatives. They were either in use (vulnerability reporting) or had been strongly proposed to be used (product liability rules).

[113] Raz, On the Functions of Law, p. 287.

widen from that of pure content. The analysis of the efficacy needs to be done on the basis of wider analysis, where the analysis of the substance of the regulation only is one part. Also other factors that influence the effects of a regulatory instrument need to be taken along.

In order to understand the efficacy of regulation it is necessary also to study the variables which intervene between the enactment of a regulatory instrument and the behaviour of the objects of regulation[114]. In the terms of regulatory studies, regulation is analysed as occurring in a regulatory space where the operation and competition of various regulatory regimes influences regulatory impact and where the effects cannot be understood without an understanding of the practices, norms and social ordering of the target population[115].

There still is a need to concentrate on the factors that are in the control or in the sphere of influence of the regulator. Otherwise the influence capability of regulation would not be possible to separate from the multitude of influencing factors. At the same time the already accepted intentionality of regulatory activity would be lost. However,

---

[114] The importance of studying the factors between the promulgation of the law and the behaviour of the public is typically raised in sociological studies of the effects of laws such as Aubert's analysis in Some Social Functions of Legislation, p. 117. Widening the analysis from that of pure content to the intervening factors implies a shift from analytical isolation to a method that utilises the empirical research results provided, e.g., by sociology of law. Instead of pure content of the norms, the analysis concentrates on the social, economic and political factors that shape the influence, similar to the sociolegal analysis as presented by Thomas Mathiesen in *Rätten i samhället*, p. 60.

[115] see, e.g., Parker et al., Introduction, p. 6-7. This regulatory space argument resembles, and mainly stems from, the responsiveness and reflexivity arguments made by a variety of scholars in sociology (e.g., Habermas, Luhmann), law (e.g., Teubner) and regulatory studies (e.g., Ayres and Braithwaite in several studies). Similar to regulatory studies, like that of Ayres and Braithwaite, *Responsive Regulation*, p. 4-5, this study does not adhere to any of the grand (general and overarching) theories that explain structures and processes of the regulation of social phenomena. The argument is more practical; it only calls for the analysis of the context of regulation, the complex abstract space where it occurs. This is necessary for the understanding of the regulation in its decentred form where the regulatory capacities, the resources and abilities of regulators, are dispersed and fragmented.

the precise isolation of the incentives provided by regulation from all other, often diverse and contradictory, conceivable factors influencing human behaviour is not possible under the classificatory principle of social functions. Every individual or a group has so many different motives and interest, and the external structures that exist in a given society that also provide incentives (such as the market in terms of prices) that no theory or empirical study has been able to establish or verify the causal link, or to propose a widely applicable general propositions on the links between the incentives provided by a regulation and the changes in individual or group behaviour[116].

Jyrki Tala acknowledges the problems in the isolation of the incentives provided by a regulatory instrument and proposes an analytical framework to trace the visible links and mechanism, and not the necessary and sufficient conditions (causal relationships), by which effects of a law reform are created and shaped[117]. The useful matrix consists of four factors in relation to which the effects of laws can be examined; the objectives, the substance (which is mainly regarded as a means of achieving the objectives), the implementation, and the reaction of the objects of a law reform[118]. Tala concludes that

---

[116] This is emphasised by Tala in *Lakien vaikutukset*, p. 354-356 in Finnish and p. 386 in the English summary. As Tala also recognises, it is a general problem in the research on human behaviour that the effects of certain specific factors cannot be established in a manner that would show it to be a necessary and sufficient condition for the behaviour. No absolute causal links can be established between an influencing factor and behaviour due to the complexity of the human motivations.

[117] Tala, *Lakien vaikutukset*, p. 354-356 in Finnish and p. 386 in the English summary. Jyrki Tala is one of the leading Finnish scholars in legislative theory and a professor in legislative research at the Faculty of Law of the University of Turku. The professorship to which Jyrki Tala acceded in the beginning of November 2005 is the first of the kind in Finland and in other Nordic countries.

[118] This matrix is based on the assumption that these factors are within the sphere of influence of the regulator, they can be influenced by the regulator, and that the four factors in particular shape the effects produced (Tala, *Lakien vaikutukset*, p. 382 in the English summary). A central limitation to the predictive capacity of this approach has to be acknowledged. Since no

each of the four factors contributes to the knowledge of the factors underpinning the creation and shaping of the effects of a law reform; however, he acknowledges the problem of differentiating between the highly interrelated and dependent four factors[119].

Even though these factors have traditionally not been central to legal scholarship, as Tala seems to be suggesting in sketching a model for the analysis of the effects of legislation, from the wider regulatory perspective this matrix is not unique. As Julia Black notes, questions concerning the influences shaping regulation, the design and deployment of regulatory tools, patterns of decision-making, strategies of monitoring and enforcement, and responses to regulation are the central issues addressed in regulatory scholarship.[120]

Jyrki Tala is sceptical about the applicability of the observations he makes about the creation and shaping of effects to statutory norms of another kind, e.g., ministerial decrees or other lower level orders by public authorities. He is of the opinion that if the way the norms are created (e.g., which actors participate) have no importance for

---

overarching theory is accepted for the basis of analysing the effects and strict causal models are abandoned, even economics with its widely used theories of human behaviour is seen just as another heuristic device among others (however good in the prediction of effects), the predictions made are more difficult to verify. Actually they could be verified only in the evaluation phase of the effects years after the enactment of the regulations.

[119] Tala, *Lakien vaikutukset*, p. 349-351 in Finnish and p. 382-384 in the English summary. Not all shaping factors can be considered, even of those which are at the margin of discretion of the regulators. They would require much more in depth and wider analysis. Only those factors that normatively belong to the instrument are analysed (i.e., are visible form the normative structure of the instrument and visible mainly for the legal analysis).

[120] Black, Law and Regulation, p. 35. In adopting the basic four factor matrix model drafted by Tala, a conscious choice of stepping into the borderline between legal and regulatory scholarship is taken. This study can be categorised as belonging into the discipline of the theory of regulation. The concentration is on the structure, norm type and content of regulation concerning the development of secure software. The focus is on the different instruments used, their incentive mechanisms and the factors that shape their influence and affect their capacity to influence behaviour.

the effects provided, which he strongly opposes, then several of the notions of the effects of a law reform could be extended also to other legal regulations[121]. This is correct as such. But this does not render the analytical four factor matrix as inapplicable for the study of the regulatory capacity of different kinds of instruments. The method, as argued in this study and widely supported by regulatory scholarship, is applicable to the analysis of all regulation. Just the observations made about the effects of a law reform with the help of the matrix are not universal and care should be taken when using them.

With objectives the analysis concentrates on the intentions of those involved in the drafting of the specific regulation. The intentions of others than just the regulators (e.g. members of the parliament and other decision-makers in legislative drafting) are of importance because typically there are a number of other actors involved with different roles (e.g., opinion leaders, those giving statements, lobbyists). In addition, the analysis of the objectives covers also the objective intentions of an instrument inside a system, detached from the intentions of those participating in the drafting. These are, for example, the general functions of an instrument (the maintenance of public order, settling of disputes and levelling of conflicting interests in the case the law) and purposes served (e.g., resolving a certain problem, the enhancement of the public interest).

The analysis of the objectives tries to answer question what is the factual situation that is supposed to come true with the regulation; the state of affairs the regulator means to bring about by particular regulatory action. When the objectives found are compared to the abstracted and assumed societal objective of achieving secure software, it becomes possible to see whether the instrument incorporates an adequate understanding of the factors affecting secure software development.

The objectives are essentially visible in the documents created during the drafting of the instrument, the statements made by the parties involved in different contexts and the studies made about the

---

[121] Tala, *Lakien vaikutukset*, p. 30-31.

instruments. However, there are also hidden intentions that are not explicitly stated. These are analysed by considering the background assumptions of different actors and the instrument inside a larger context and a part of a system. The background assumptions are made visible by using a hermeneutic approach[122].

The analysis of the objectives is essentially speculative in nature. First of all, the instruments have not been explicitly stipulated, i.e., they are constructions of practice instead of explicit drafting[123]. This hinders the analysis of the intentions and objectives of specific regulators since they can vary a lot. In these cases the regulators are considered, where feasible, as groups that have at least somewhat similar incentive structures and can be expected to have similar types of objectives. The huge variety of differing objectives inside these groups still has to be acknowledged even though analysis cannot make them visible.

Also the general nature of many of the regulations, i.e., they are drafted to fit a huge variety of situations, like in the case of contractual and non-contractual liability doctrines, instead of reflecting a single socially constructed area of concern such as the software development, emphasises the speculative nature of the analysis of objectives[124]. Many of such general regulations cut across number of social contexts and in each case the contextual facts will trigger a unique set of concerns and goals. Thus, the analysis of the objectives of such general regulations in the social context of secure software development is

---

[122] For a short overview of hermeneutics in law, see Morawetz, Law and Literature, p. 456-457. This is the core of the legal positivist method of inquiry as best represented by H.L.A. Hart in jurisprudence. Note that I am not interested in the discussions about the generally applicable concept of law or regulation (criteria of legality). Instead, I am only interested in the ways regulation affects behaviour. This is the second essential question of legal positivism (normativity of law).

[123] This has been raised in relation to tort doctrines by Jane Stapleton in Regulating Torts, p. 132, where she critically evaluates Hugh Collins regulatory approach to contract law and its applicability to tort law.

[124] Stapleton in Regulating Torts, p. 133, raises this in relation to tort doctrines.

essentially a speculative work, especially when no practical cases exist where the contextual factors have been considered[125].

The actual objectives of the regulators in real implementations or the effects of specific instruments in use are not analysed in this study. The concentration is on general ideal types of instruments and no empirical research in conducted. The policy papers, the drafting documents etc. provide information of the general objectives of the regulators and the purposes of the instrument used, but no interviews have been made[126]. As Tala emphasises, the limitations of the information concerning the objectives has to be recognised; the sources that give historic or empirical information can be used to depict the objectives, and this information is useful in the analysis of the regulatory capacity, but no reliable, final or complete knowledge can be achieved[127]. This is a clear lacking in the study which hopefully will be corrected in further studies.

Since this study strives for an understanding of the possibilities of regulation in solving problems in secure software development,

---

[125] What this means is that the objectives gleamed out in such cases are the mental work of the analyst, instead of explicitly being stated or even considered by the regulators. However, this concern the whole analysis of the objectives since they are gleamed out from the background assumptions of the regulators and only in rare cases can explicit statements especially in relation to secure software development be found. Even in cases where the objectives are explicitly stated they need to be compared to the incentive structure and background assumption of the regulators in order to test whether they are mere lip-service or rhetoric. The objective intentions of the instruments inside a system, i.e., their inner logic which in the case of legal regulation typically are the creations of scholars, are also useful in this.

[126] This essentially is, and can be, just a preliminary analysis. Some of the instruments (especially vulnerability reporting) are so new and subject to rapid changes that a sufficient understanding neither of the influence mechanisms nor the related objectives has not yet been gained. This makes the writings and other representations of the attitudes of those participating in the regulatory processes important. This is why writings in newspapers and discussion lists are relevant in relation to the new and informal types of regulations like vulnerability reporting.

[127] Tala, *Lakien vaikutukset*, p. 102.

rather than an ex post empirical analysis of the effects, also the objectives of regulators are more or less theoretical and result of the understanding of the context of secure software development. This emphasis the uncertainty and incompleteness of this knowledge. Despite of this, to the degree the objectives can be made explicit, this information is useful in understanding and analysing regulatory capacity. The speculative nature of the information just has to be acknowledged, and the need for empirical verification after information on the implementation of the instruments can be gathered from actual implementations[128].

The analysis of the substance of regulation concentrates on the content of rules and the way it shapes the effects regulation can have. The internal view of law, the descriptive and prescriptive legal scholarship by concentrating on the finding of the best justifiable meaning of a regulatory instrument, is useful in the analysis of the content of a norm. It gives information on the ways the objects are supposed to behave and thus of the intended effects of the regulation[129]. However, an instrument can also have unintended effects and effects that appear outside the scope of the norm. Also policy

---

[128] An analysis of specific implementation in certain jurisdictions needs to acknowledge the variations of objectives between international actors, jurisdictions and areas with different regulatory styles.

[129] Even though legal scholarship, the internal view of law, forms a part of the method by which the substance of specific instruments are analysed, the purpose in this study is not to find the best justifiable meanings of a regulatory instrument. The purpose of the study is not to give recommended interpretations of specific norms or the systematisation of norms according to the traditions of a specific jurisdiction. No normative arguments about the correct interpretation of the substance of the norms are provided. The instruments are analysed as abstract models (ideal type analytical tools) starting from the dispositive rules (the standardised rules applicable to situations where nothing else has been agreed on) and the basic elements of the instruments. The uncertainties in the interpretations of the substance of the instruments and their variations between contexts (e.g., contracts that typically deviate from the dispositive rules) and jurisdictions is acknowledged and the effects of this uncertainty are considered, but no normative arguments about the right interpretation of the substances are provided.

analysis type of study (traditional toolkit analysis) gives information on the assumed effects of the regulation because the tool choice is partially based on the assumptions how people react to the specific types of regulation[130]. Policy analysis typically considers different regulatory options, their differences, benefits and disadvantages.

As was already noted above, the analysis of the substance, even when accompanied with the objectives, does not tell about the actual effect of a regulatory instrument. Only the intended effects and at best just potential effects are found. Between the substance of regulation and the effects created by the regulation exists a variety if different factors and mechanisms that shape the effects. The expansion of the analysis from that of the substance (the content) and the intended effects is necessary because the effects caused by a regulatory instrument, even if it is normatively binding as law is, does not translate into behaviour automatically, mechanically or fully in a way it is intended. The regulation can also provide incentives that were neither foreseen nor intended by the regulator and they can appear outside the target subject matter. The guidance provided by the regulation is altered when the objects of regulation apply it. Their choices and decisions are influenced also by other factors than just norm in question and the knowledge of its correct substance (the standard of conduct it provides).[131]

When the essence of regulation is influencing or altering behaviour, it emphasises the relational nature of regulation. It is not just imposed from above and the effects are not in direct causal relationship with the regulation. Instead, regulation is just one factor in the governance of an issue and questions of how the regulation is implemented and how the objects react become important. This highlights the importance of the actors under the scope of a regulation, the structure and the operational environment they form, for the shape of the

---

[130] These assumptions are largely based on welfare economics but has developed further as depicted in basic textbooks on policy analysis, such as that of Weimer and Vining, *Policy Analysis*.

[131] Tala, *Lakien vaikutukset*, p. 35-36.

effects of a regulatory instrument. The implementation (the actors putting the rules into practices) and the reactions of the objects form a mechanism that mediates the substance together with the objectives and the effects.[132]

With implementation the interaction between the actors doing the necessary task required for the norm to be effective and the objects of regulation are of especial interest. Among the question answered are: how many mediators there are, who is responsible and pays for the implementation, are the monetary and human resources sufficient (e.g. the number of persons and the level of expertise). In the analysis of the reactions of the objects of regulation, the first task is to identify to whom the specific regulatory instrument is addressed – whose behaviour is being directed. The starting point is that, in addition to the preferences and incentive structures of the implementers, the effects are shaped also by the attitudes and beliefs of the objects of regulation.

The analysis of the reactions of the objects has been found to be more fruitful when starting from the different actors, their options and decision-making contexts rather than from the regulation itself[133]. Especially considerations of the decision-making models of different actors, the role played by them (e.g. private individual or consumer, public organisation, SME or big commercial actor), the constraints to their rationality and capabilities (e.g. level of expertise and knowledge) and their resources are relevant. The likelihood of opportunism or of voluntary compliance, which is affected by the incentive structure, is also relevant.[134]

---

[132] Tala, *Lakien vaikutukset*, p. 174 and 177-178.

[133] Tala, *Lakien vaikutukset*, p. 323.

[134] Especially this analysis demands a deeper understanding of the characteristics of secure software development and the incentive structures present. Such an analysis is performed largely when answering the first supporting research question and analysing the problems in secure software development in chapter 2.

A special challenge for the information gathering in this study is that the analysis is not made in relation a specific regulatory action in a certain time frame and in a certain context or even a jurisdiction[135]. Instead the analysis concerns the abstract ideal types of instruments, not a specific realised regulatory project or one being carried out in practice in a specific context. The analysis is based on theoretical considerations and evidence provided by others[136].

An additional problem is that wide empirical knowledge of the instruments in general and their effects have not been gathered systematically nor critically evaluated even by others. For this reason, the analysis involves a huge amount of assumptions and the results are just indicative despite the assumptions being widely held and theoretically studied extensively[137]. Thus the analysis is based on a prediction of the effects of regulation, not an empirical verification conducted after the enactment of the specific instrument. This is due to the study concentrating only the ideal types of instruments, not a specific realised regulatory measure inside a jurisdiction or a specific situation, or one being carried out in practice. This leaves the analysis of the possible effects of regulation on secure software development exposed to rival interpretations. Hard evidence is scarce.

Because only abstract ideal types of instruments are analysed, the information collection is restricted. No empirical research is conducted in this study nor is it feasible in a wider context. An empirical analysis

---

[135] The abstraction from specific regulatory systems has been the starting point of regulatory theory. For example, Peter Noll did not bind his arguments into a specific legal system as pointed out by Jan Hellner in *Lagstiftning inom förmögenhetsrätten*, p. 148.

[136] Some of the instruments analysed are so new and still evolving that information about them, even of their structure, is just being gathered. No hard facts can be stated even of the form some of the instruments take. The evidence gathered is patchy and susceptible to rival interpretations.

[137] The lack of detailed, systematic empirical studies is a severe hindrance. The casual observations about the reaction of the objects made on the basis of immediate appearances that dominate many discussions of the public's behaviour can be deceiving. An unfounded belief in having understood the public is a serious barrier to acquiring a genuine understanding.

of the effects created by regulation for a specific issue (analysis of the effects of a variety of regulations on a single issue instead of evaluation of a specific regulation on a variety of issues) is not feasible in any sufficient amount of detail because it would require a particular implementations in a specific context, organisation and jurisdiction, and a longer lapse of time after the enactment of the regulations[138]. Such information would be difficult to generalise and of little use for different regulators. In addition, an empirical analysis even of a specific instrument in a single category would require significant research resources. Just an empirical analysis of a specific instrument in a category would require resource not possible for a single researcher.

Neither can most of the information collection and analysis methods used in the regulatory impact analysis of specific instrument implementations in certain contexts (public consultation, engineering studies, survey design approaches or econometric approaches) be used[139]. When such analyses have been conducted in certain implementations, they are utilised with care. For example, public consultations performed in specific implementations in certain jurisdiction will be used, even though they are rare in relation to issues such as information security[140]. However, the generalisation cannot

---

[138] It has to be made clear the no empirical analysis neither of the implementation nor of the reactions of the objects is done in this study or could even have been possible. As Jyrki Tala points out in *Lakien vaikutukset*, p. 237, the implementation analysis can be done only years after the specific regulation has been enacted due to the changing circumstances and the learning that occurs with time.

[139] Of these types of information gathering techniques used to determine the behavioural response of the regulated entities, see OECD, *Regulatory Impact Analysis*, p. 247-252.

[140] The restriction of such public consultations must be beard in mind; these forms of information gathering are inevitably weighted in favour of those active and knowledgeable participants who provide information. More of the restrictions of public consultations see OECD, *Regulatory Impact Analysis*, p. 247-248).

be done directly for every jurisdiction and context[141]. Even though the analysis concentrates on regulatory tools that seem to be more neutral as far as legal and regulatory traditions and culture are concerned than do institutions, where historical, social and economic conditions play an important role in their formation, there still are significant variations[142]. Without detailed considerations of the influences of the differences of regulatory systems and styles between countries and wider supranational systems, justice is not done to the significance of local differences and the variety of regulatory styles in comparative legal systems.[143]

However, a laborious comparative analysis considering the contextual and local nuances is not possible in any sufficient amount of detail with a variety of instruments plainly due to the complexity of such analysis. The number of variables to be considered would increase tremendously. Rooting the study into the peculiarities of a single regulatory system is neither justified since the results of an analysis of the capacities of specific regulatory instruments to influence secure software development being applied in a certain jurisdiction and a context would be difficult to generalise for the exact same

---

[141] The generalisation of the arguments in regulatory theory is hampered by the effects of national legal system, such as those made explicit by Daintith, Law as Policy Instrument, p. 35, in an study about the legal implementation of economic policy instruments: the formal and explicit constitutional requirements of the systems, the regulatory style in use and the substance of the specific national regulatory system in general. However, it has to be noted that the problems in generalising the regulatory theory arguments between jurisdictions and specific contexts does not render the use of international material impossible. It just means that the material should be used with care and the possible national nuances recognised.

[142] This has been pointed out by Attila Harmathy in The Influence of Legal Systems on Modes of Implementation of Economic Policy, p. 255, in analysing of the influence of legal systems on the use of different regulatory instruments in the legal implementation of economic policy. Different legal systems respond to their cultural context and local economic conventions, so that their regulation assumes different forms and institutions.

[143] It is important to note that the analysis is based on a European perspective and the examples are mainly taken from the EU level, and from Nordic or national Finnish regulation.

reasons. However, while endeavouring to generalise, the significance of the local differences have without a doubt been downsized.

Even though this qualification limits the usefulness of the insights gathered about the effects of regulation in that the utilisation of the results requires care in recognising and considering of the possible national and contextual nuances, the more abstract analysis is justified on the grounds that the possibilities of several instruments can be analysed to solve a problem that is essentially international in nature. This limitation follows from the purpose of this study to find reasons and solutions for the current problems in secure software development; rooting the analysis into a specific regulatory system would limit the available solutions unnecessarily.

The matrix of Tala needs to be added with the identification of the regulators in the decentred understanding of regulation. It is often not obvious who is doing the regulating and it is not sufficient just to analyse the regulators in relation to the objectives. Vice versa, the objectives and conflicts between them are more easily seen when the different regulators are made explicit. Who the regulators are and in what institutional structures they operate directly affects, among other things, the legitimacy of the regulation in the eyes of the addressees.

In the identification of the regulators in the secure software development, which is done in heading 3.5, an iterative process of searching and specifying is used. The starting point are the general categories identified in previous regulatory studies. This is followed by an ad hoc collection of actors. The previous knowledge of the author is amended by literature analysis performed in order to answer the first research question and by general observation of the subject matter in the news. This ad hoc collection makes visible certain types of actors that are convenient to combine into a category. At the same time it is possible to omit less important actors.

> "… *an overwhelming majority of security vulnerabilities are caused by 'buggy' code.*"
>
> (Schneider FB (ed., 1999) *Trust in Cyberspace*, Committee on Information Systems Trustworthiness, Computer Science and Telecommunications Board, National Research Council, National Academy Press, Washington, D.C., p. 110, http://bob.nap.edu/html/trust/ [8.3.2006])

# 2 Understanding secure software development[144]

The main method of securing information systems and software is adding security afterwards; they still are not developed with security in mind[145]. This applies to both aspects of secure software and information system development: the quality aspects of security (reducing vulnerabilities that could pose a risk to security) and the security features that can be seen as requirements of an information system or a software and should be made part of the system according to user needs. The former prevents threats stemming from the system itself (by reducing vulnerabilities in the system) and the latter prevents

---

[144] The first part of this chapter discussing the influence of the network economic environment has been presented in the 37th Hawaii International Conference on System Sciences, January 5-8, 2004, Big Island of Hawaii. The paper (Råman 2004) has been published in the proceedings. Special thanks to the anonymous referees for the excellent comments and suggestions for amendments and to the participants of the Minitrack on Information Systems Security Management.

[145] This argument has been raised, e.g., by Premkumar Devanbu and Stuart Stubblebine in Software Engineering for Security, p. 228-229, while mapping the future of software engineering in 2000. It is still valid despite the relatively old and expanding legal obligations to consider security in information systems development such as the section 18 of the Finnish Act on the Openness of Government Activities (621/1999) that obligates governmental organisations to take security into consideration when developing information systems. Similar provisions have been enacted for many sectors. In the following, the reasons for this are considered.

threats coming from outside the system, e.g., accidental or deliberate abuses of vulnerabilities in the system or persons using it.

The quality aspect of security is dominated by the "penetrate-and-patch" approach[146]. Vulnerability avoidance is still overlooked, and security becomes an issue only after a published security breach. This is so despite of the evidence showing that the practice of waiting until the end of the development cycle to deal with the vulnerabilities is wasteful[147].

In software engineering, security requirements are seen as complementary to the normal, or functional, requirements of a system (such as the features that the customer would require). The favoured methods for requirements engineering typically do not even include security concerns as an integral part of the process. Although some security concerns are addressed during the requirements engineering stage, most requirements come to light only after functional requirements have been satisfied. As a result, security requirements are added as an afterthought to the standard (functional)

---

[146] This has been raised especially by practitioners of secure software development, such as John Viega and Gary McGraw in *Building Secure Software*, p. 15-16.

[147] Researchers have provided evidence of economic benefits from early attention to security vulnerabilities. After looking at the cost and benefits of increased security on application development, Soo Hoo et al., Tangible ROI through Secure Software Engineering, p. 3, concluded in 2001 that catching the already known or reasonably easy to find security vulnerabilities in the design phase was more cost effective than in implementation, which is more cost effective than in testing, and so on. This has been pointed out also by Viega and McGraw in *Building Secure Software*, p. 16. Building security (in the sense of fewer vulnerabilities) into applications from the start improves reliability, avoids potentially embarrassing and costly incidents, and ultimately saves money. The basic economic rationale has been explained already in 1981 by Barry Boehm in *Software Engineering Economics*, p. 39-41, in relation to non-security related errors in general. As Blackburn et al., show in Improving Speed and Productivity of Software Development, p. 884, this kind of attention to quality (customer expectations and requirements) early in the life cycle of a project (at the requirements analysis and specification level) leads to defect detection and avoidance that both increases productivity and improves speed to market.

requirements[148]. This is the case despite the reasonably wide agreement that security requirements (e.g., confidentiality, integrity and availability) and other security features (e.g., access rights, security classifications) should be defined into the system from the beginning and despite the efforts to integrate security design into the development processes[149].

One explanation for the low quality of software, when compared to other practice (engineering) and research fields, stems from the immaturity of software engineering and information systems disciplines. The methods for development are constantly evolving (novel methods arise every now and then, and are modified by practitioners to fit different situations), and they do not consider information security design issues, which are relegated to separate secure software and information systems development (developmental duality)[150]. In addition, one of the important reasons for the inadequacy of the space in which software designers today operate is "that the market structures within which software development

---

[148] Devanbu and Stubblebine, Software Engineering for Security, p. 227-228.

[149] The widely applauded new OECD 2002 *Guidelines for the Security of Information Systems and Networks*, p. 8, recognize this in arguing in favour of a culture of security – that is, a focus on security in the development of information systems and networks and the adoption of new ways of thinking and behaving when using and interacting within information systems and networks. The Guidelines signal – as a consensus of several international professionals a clear break with a time when secure design and use of networks and systems were too often afterthoughts. Note that the guidelines are directed at the developers of network and information systems. They are customers of the packaged software vendors. However, this does not diminish the weight of evidence provided of the change of paradigm.

[150] As originally noted by Richard Baskerville already in 1992, The Developmental Duality of Information Systems Security, p. 6-9, the separation of normal information system development from the information security process creates problems such as restrictions on proper system functions, higher costs and shorter system lifespan.

occurs are still primitive in comparison to those supporting other industries"[151].

The widely recognised lack of experts in software development that have knowledge in security is also an explanation for this contradictory behaviour, i.e., security is not taken seriously in the development phase, even though there is evidence of extensive benefits and agreement on the importance of doing so. Vendors simply lack expertise in secure development. Secure software development has not been, and still largely is not, part of the curriculum in most educational institutions teaching software developers. Not even in the country producing most software, i.e., USA[152]. With the realisation of the central role of the security of the software in the current infrastructures, for the commerce, and for the society at large, this has become a crucial lacking. As Bruce Schneier notes in the foreword for one of the first English language textbook on secure software development: "We need better education. Programmers must learn how to build security into their software design, and how to write code securely"[153].

However, these can only be partial explanations. While the state of methods for development is still rather inadequate and experts are a scarce resource, there are well-known technical and procedural ways to prevent at least the widely known vulnerabilities and methods to integrate security in development processes[154]. Thus, given enough

---

[151] This is what Barry Boehm and Kevin Sullivan argue in Software Economics, p. 329, while mapping the future of software economics as part of software engineering in 2000.

[152] National Cyber Security Partnership, *Improving Security Across the Software Development LifeCycle*, p. 3-4.

[153] Schneier, Foreword, p. xx.

[154] For a somewhat comprehensive compilation of technical information security standards from the United States in 2004, see NCSP, *Technical Standards and Common Criteria*, Appendix C. A roadmap for the ICT security standards is provided by the International Telecommunications Union's Telecommunication Standardization Sector (ITU-T) where both the standardization organisations, approved standards and standards under development and new proposed standards are presented. See the ITU-T Study Group 17 ICT

demand from customers, vendors could supply improved quality and security. Also more experts in secure software development are likely to arise if knowledge in security becomes a competitive advantage.

It is not that developers are incapable of producing software with less vulnerability or including security features into systems; they just are not sufficiently motivated to do so. Beyond the intrinsic difficulty of writing defect-free software[155], there are constraints on the development of secure software that result from the business environment: the constraints for the COTS software business model largely derive from schedule and budget[156].

In this chapter I try explain this contradictory behaviour by further analysing the influence of the constraints from the business environment on secure software development. The effort is to understand the reasons why security has typically not been considered in the development phase[157]. In order to be able to understand the

---

Security Standards Roadmap v1.0, November 2005, at http://www.itu.int/ITU-T/studygroups/com17/ict/index.html [updated 25.1.2006, visited 1.2.2006]. In addition, several software process improvement methods have been proposed and many automated tools also exist to support the software developer in security work.

[155] It is notoriously difficult to write defect-free software; especially large and complex software of our days. Also software developers have to live with the fact that every product contains errors, some of which may not materialise until a particular and perhaps unrepeatable, or even malicious set of circumstances occur. However, this does not prevent software from being 'good enough' for the specific situation and customer needs. The expectation is not, it cannot be a flawless product.

[156] To state the obvious: it is not feasible to invest infinite resources on security development since it is just one requirement and a quality factor that needs to be fulfilled. Secure development is essentially about balancing costs and benefits both in terms of schedule and budget.

[157] Since the analysis relies on literary sources and concentrates especially on academic research, there necessarily is a time lag in relation to the present methods of secure software development. In this sense, the analysis stems from an abstract situation which was largely present at the turn of the millennium where the quality aspect of security had not been of much interest and no major effort had been put into improving the security of software. The underlying argument is that the increasing interest in secure software

importance and the extent of the schedule and budget constraints for market-driven development a short look at the economics under which market driven development operates – network or information economics – is first needed[158].

## 2.1 The network economic environment

Software as an information good has an unusual cost structure[159]. An information product is typically expensive to produce, but very cheap to reproduce. The fixed cost related to producing the first copy are not only fixed but also often largely sunk: fixed costs are not recoverable if production is halted[160]. Also variable costs are typically

---

development during the last couple of years together with the consequent improvements is partly due to the regulatory pressure; i.e., software vendors desire to improve the security of software has partly been due to the increased threat of further regulatory action. This development has not, however, been enough to change the state of security in commercial-of-the-self software.

[158] The following presentation is heavily based on the influential first comprehensive presentation of the economics of the "new" business that was subject to lively debate in the later half of the 1990s by Carl Shapiro and Hal Varian, *Information Rules. A Strategic Guide to the Network Economy*, Harvard Business School Press, Boston, Massachusetts, 1999. The basic economic structure has not changed since even though the business has. Those familiar with software economics issues are welcome to proceed to the following section.

[159] Remember that the concentration is on COTS software. The economic basis of open source software is not considered in this study.

[160] This is a crude generalisation – the costs are not completely sunk. When software undergoes a series of releases that stem from modifying the existing code and care is put into the design and coding, then less code needs rewriting. Design documentation makes changes easier and if the same people make the modifications then their experience makes future development cheaper. Software engineering methods and processes such as reuse, not just of software code but also the knowledge gained from past projects, of which see, e.g., Ellmer et al., Process Model Reuse to Promote Organizational Learning in Software Development, p. 21-26) are intended to allow developers to gain from earlier development work and making it possible to recover at least some of the cost of developing stable code. This has been pointed out, e.g., by Richard Botting in On the Economics of Mass-Marketed Software, p. 467,

small: the cost of producing an additional copy typically does not increase, even if a great many copies are made.[161]

Substantial supply-side economies of scale (lower unit costs by being larger) are gained with this 'high fixed cost of development and low marginal cost of producing subsequent copies' feature typical in software and information markets (similar to most industries). But demand-side economics of scale (taking advantage of positive relationship between popularity and value) is the norm in information industries. The combination of both demand- and supply-side economics of scale makes the information industries different: growth on the demand side both reduces cost on the supply side and makes the product more attractive to other users; this accelerates the growth in demand even more, resulting in especially strong positive feedback.[162]

Networks of compatible users generate *network effects* (i.e. value of the software to the individual user depends on how many other user there are for the same software product) that in turn give rise to positive feedback[163]: as more users deploy the same software, the more communication partners there are to share files and tips with and the more encouragement there is for software houses to devote more resources to developing compatible software. This drives the potential users of software to buy the product they believe to become the dominating one and to keep its position. Thereby they get the most value for their money.

---

while discussing the economic forces acting on the COTS software producer in 1997.

[161] Shapiro and Varian, *Information Rules*, p. 3 and 21; Rajala et al., Software Business Models, p. 21.

[162] Shapiro and Varian, *Information Rules*, p. 21 and 179-182.

[163] Shapiro and Varian, *Information Rules*, pp. 14 and 183. Stan Liebowitz and Stephen Margolis argue that the term "network effects" should be applied to markets with increasing returns to scale and the term "network externalities" reserved for markets in which increasing returns create sub-optimal conditions (Liebowitz and Margolis, Network Externality, p. 133; similarly in Lemley and McGowan, *Legal Implications of Network Economic Effects*, p. 5). So, network effects should not properly be called network externalities unless the participants in the market *fail* to internalize these effects.

Network markets can be viewed as falling on a continuum that may roughly be divided into actual (direct) networks, virtual (indirect) networks, and simple positive-feedback phenomena[164]. Computer software has been seen as a paradigm example of virtual networks - a good that provides inherent value to consumers that increases with the number of additional users or identical and/or interoperable goods[165].

In addition to horizontal technological compatibility, software is subject to increasing returns based on positive feedback from the market in the form of complementary goods. Software developers will write more application programs for an operating system with a bigger market share because that operating system will provide the biggest market for applications programs. Conversely, the availability of a broader array of application programs will reinforce the popularity of an operating system. This makes investment in application programs compatible with that system more desirable than investment in programs compatible with less popular systems. Similarly, firms that adopt relatively popular software (not just an operating system) will likely incur lower costs in training employees and will find it easier to hire productive temporary help than will firms using unpopular

---

[164] This classification has been made by two U.S. based legal scholars Mark Lemley and David McGowan in *Legal Implications of Network Economic Effects*, p. 13. The essential criterion for locating a good along this continuum is the degree to which the good provides inherent value to a consumer apart from any network characteristics. The greater the inherent value of the good relative to any value added by additional consumers, the less significant the network effect.

[165] Lemley and McGowan, *Legal Implications of Network Economic Effects*, p. 18. Virtual network goods need not be linked to a common system, as the constituents of a communications network (actual network) are: very strong positive feedback effects tied to functional compatibility are sufficient. Unlike actual networks (e.g., telephones and fax machines where the entire value of the product lies in facilitating interactions between users and the benefit to a purchaser is access to other purchasers), goods that constitute virtual networks (e.g., an operating system or an application program) allow even a single user to perform a variety of tasks regardless of other users of the software. However, the value of a given software product grows considerably as the number of additional purchasers increases (e.g., easier file sharing, less need for retraining and thus more competent employees available). (Idem.)

software.[166]

One of the most striking consequences of network effects is their impact on the nature of competition between sellers of products embodying different, incompatible standards. When firms compete for a market where there is strong positive feedback, only one will emerge as the winner. As the installed base of users grows, more and more users find adoption worthwhile and the product eventually achieves critical mass (a large enough customer base) and takes over the market[167]. As a consequence, growth becomes a strategic imperative, and not just to achieve the usual supply side economics of scale but also to achieve the demand side economics of scale generated by network effects. Obtaining critical mass becomes the key challenge, after which the market is considered to build itself[168]. In its most extreme form, positive feedback can lead to a winner-take-all market, in which a single firm or technology vanquishes all others[169].

This tendency of network markets to tip leads to particularly intense competition early in the market's existence. As networks take time to build up to critical mass (i.e., become widespread enough to be economically viable), producers that are sufficiently ahead of the competition (in both time and appeal to the market) with a new product or application will be able to acquire the necessary critical mass to exploit economics of scale. Accordingly, the best way to

---

[166] Lemley and McGowan, *Legal Implications of Network Economic Effects*, p. 19. Note that the strength of network effects will vary depending on the type of software in question. Network effects will be materially greater for operating systems software than for applications programs. (idem.)

[167] Such markets are called "tippy", meaning that it can tip in favour of one player or another. It is unlikely that all will survive (Shapiro and Varian, *Information Rules*, p. 176). Whether a market tips or not depends on the balance between economies of scale (either in demand or supply side of the market) and variety - with strong scale economies, the market is likely to be tippy, and if different users have highly distinct needs, the market is less likely to tip (Idem p. 188).

[168] Shapiro and Varian, *Information Rules*, p. 14.

[169] Shapiro and Varian, *Information Rules*, p. 177.

secure market leadership in the presence of the economics of scale typical of information industries is through an early presence in the market: simply being first to market can generate both differentiation and cost advantages[170]. The key is to convert the timing advantage into a more lasting edge by building an installed base of users[171].

## 2.2 Time-to-market and security

Network effects have considerable influence on the behaviour of COTS software producers in particular: short time-to-market to exploit the first-mover advantage is crucial to establish a software product in the market and to profit from the network effects. Incentives to be the first on the market and to establish one's own products as de facto standards are very high. Studies in software development repeat the economic rationality that in market-driven development the primary goal is *time-to-market*[172]. Time-to-market is crucial not only for a new system but also for new product features or concepts in existing systems. Nor does time-to-market constrain only the initial release: market leaders must keep developing advanced features, bug fixes, and performance improvements in order to keep old customers satisfied and to win new ones[173]. In fact, competition for best time-to-market is perpetual.

*Short development cycle* due to time-to-market pressures is a general phenomenon in market-driven development, especially in the

---

[170] Shapiro and Varian, *Information Rules*, p. 30 and 146.

[171] Note that being first to market is not necessarily decisive in the long run, even though it usually helps as Shapiro and Varian note in *Information Rules*, p. 181-182.

[172] See, e.g., Carlshamre, *A Usability Perspective on Requirements Engineering*, p. 58; Natt och Dag, *Elicitation and Management of User Requirements in Market-Driven Software Development*, p. 17; Potts, Invented Requirements and Imagined Customers; Carlshamre and Regnell, Requirements Lifecycle Management and Release Planning in Market-Driven Requirements Engineering Processes, p. 961.

[173] Sawyer et al., Improving Market-Driven RE Processes, p. 223; Baskerville et al., How Internet Software Companies Negotiate Quality, p. 52

development of mass-marketed software (including software embedded in consumer electronics and telecommunications equipment)[174] and Internet software. Already in 1998 the development cycles have compressed from 24-36 months to 12-18 months for non-Internet related companies and even to 3-6 months for companies involved in eCommerce and creating and maintaining Web portals[175].

The high incentives to be first to market and to establish one's own products as de facto standards means that software development practice easily becomes distorted to maximize functionality and minimize development time, with little attention paid to other qualities (especially non-functional requirements such as security and safety)[176] or project goals that conflict with functionality and time-to-market (fundamentally so in the case of security because more security limits functionality)[177]. Anecdotal evidence shows that companies may compress their quality assurance (QA) practices when the priority is a shorter development cycle[178]. A shorter development cycle results in reduced quality and security, especially if ad hoc processes are used – as they typically are – in the prioritisation and cost/impact

---

[174] This was subject to much interest in the late 1990's and has been reported especially by Cusumano and Shelby, *Microsoft Secrets*, p. 15; Blackburn et al., Improving Speed and Productivity of Software Development, p. 875; Kuvaja et al., Specific Requirements for Assessing Embedded Product Development, p. 74; Botting, On the Economics of Mass-Marketed Software, p. 327.

[175] Cusumano and Yoffie, *Competing on Internet Time*, p. 298-299. Shortening development cycles in Internet software companies has been reported also by MacCormack in Product Development Practices that Work and by MacCormack et al., Developing Products on "Internet Time", p. 144-145, and in relation to application and smaller niche software developers in Baskerville et al., How Internet Software Companies Negotiate Quality, p. 51, and Baskerville and Pries-Heje, Racing the e-bomb, p. 55.

[176] This has been emphasised by Fred Schneider in *Trust in Cyberspace*, p. 67 and 72.

[177] Viega and McGraw, *Building Secure Software*, p. 26 and 35 The complexity of software engineering under conflicting requirements has been emphasised, e.g., by Barry Boehm in *Software Engineering Economics*, p. 20-21.

[178] Baskerville et al., How Internet Software Companies Negotiate Quality, p. 51; Wheterbe and Frolick, Cycle Time Reduction.

assessment of sometimes conflicting requirements such as quality/security costs and time-to-market[179].

For example, despite the wide acknowledgement of the advantages of good requirements specification early in the development phase, the documentation and maintenance of requirements tend to be sketchy. Requirements that are proposed, invented, or designed are communicated within the development organisation to a large extent by word of mouth; they are not elicited or played back to the customer since in market-driven development users (referred to as customers) are typically unknown during the development phase (there may not even be a user until the first release of the product)[180]. Even though there is increasing research on requirements engineering due to the wide recognition of the centrality of requirements specification to the whole development process already in the late 1990s[181] and improvements have been made, many of the general challenges in traditional requirements engineering are adopted by the market-driven development organisation (e.g. requirements are erroneous, errors are detected late, and ambiguities are difficult to resolve)[182].

---

[179] Sawyer et al., Improving Market-Driven RE Processes, p. 227.

[180] The way software-consuming organisations interact with software producers has changed due to the increased product attention. Traditionally, early and close links between users and developers have been considered critical. Today, software consumers and producers use a variety of intermediated means to communicate their needs to developers. For example, packaged software developers build to requirements gleaned from a variety of sources, including help-desk call-log analysis, market research, product reviews, and user groups, of which direct customer contact is one of the least likely means. (Sawyer, 2001, p. 100, see also Keil and Carmel, 1995, p. 33–44) This is one of the major dilemmas: the challenges of developing software for larger markets is to satisfy the end user although contact with the end user is limited (Natt och Dag, *Elicitation and Management of User Requirements in Market-Driven Software Development*, p. 19).

[181] See, e.g., Blackburn et al., Improving Speed and Productivity of Software Development, p. 884; Schneider, *Trust in Cyberspace*, pp. 68-74

[182] Natt och Dag, Elicitation and Management of User Requirements in Market-Driven Software Development, p. 26. This is often because time pressures change the way requirements are approached. Traditionally, requirements analysis has been based on the assumption that a large company

An important consequence of the shortening development cycle (due to first-to-market being essential) is the tendency towards *release-oriented development*[183]. Constantly striving to be ahead of competitors, the market-driven development company frequently delivers new and improved releases (containing bug fixes and new features) of software products. The development resources – especially time-to-market, but also budget – are typically fixed[184], at the expense of lower-priority requirements. Each feature is examined to determine whether its inclusion in the product is necessary for the product to be competitive in the marketplace.

Generally, those features with direct customer appeal win; subtle, hard-to-demonstrate and pervasive properties – such as security – tend to be rejected[185]. They are put of from one release to another

---

or institution orders a large software system from either an external vendor or internal IT department, using a requirements specification as a contract. However, in market-driven development there is no contractual situation, and requirements specifications are rarely written (Carlshamre, *A Usability Perspective on Requirements Engineering*, p. 57-58; Natt och Dag, *Elicitation and Management of User Requirements in Market-Driven Software Development*, p. 21). Dahlstedt et al. in Market-Driven Requirements Engineering Processes for Software Products – a Report on Current Practices, partially verify many of the characteristics of market driven requirements engineering in a first stage industrial survey of small and fairly new market-driven development companies in the Swedish software industry.

[183] This has been reported for market-driven software development in general (see, e.g., Natt och Dag, *Elicitation and Management of User Requirements in Market-Driven Software Development*, p. 17; Potts, Invented Requirements and Imagined Customers; Karlsson and Ryan, A Cost-Value Approach for Prioritizing Requirements, p. 68) as well as software for electronic commerce and for web portals (see, e.g., Cusumano and Yoffie, *Competing on Internet Time*, p. 14, and 224-234; Baskerville et al., How Internet Software Companies Negotiate Quality, p. 52). In interviewing companies using Internet speed development techniques Baskerville and Pries-Heje, Racing the e-bomb, p. 56-57, found that vague requirements continue throughout projects. This is one cause for release orientation.

[184] This argument has been raised, e.g., by Ross Anderson, a Professor of Security Engineering at the Computer Laboratory of the University of Cambridge, in a seminal paper combining economics and information security in 2001, *Why Information Security is Hard* p. 2.

[185] Schneider, *Trust in Cyberspace*, p. 194.

in order to meet the release date. This feature or requirement slippage should not come at the expense of quality and security, even though this easily is the case[186]. Security tends to be ignored because it would require that more time and money be put into development work that is beneficial only in the long run (no immediate returns). Time-to-market considerations discourage the inclusion of security features and encourage the postponement of security to later releases – if they are considered at all.

A release orientation is necessary for a start-up company since it cannot generate revenue before it produces functionality, even where this is somewhat unreliable. The code of the first release lacks security because any issues that can be postponed, including quality and security, are disregarded[187]. For an existing (or even dominating) firm, the frequency and timing of new versions and upgrades is a way to control the length of the cycle of customer lock-in into one's products as economic theory explains[188]. It is necessary to produce new releases and upgrades to prevent aggregate customer lock-in from getting too low at any point of in time, because the optimal time for a competitor to enter the market and to attack the installed customer base of the existing firm is when the aggregate customer lock-in is low[189].

However, the quality aspect of security is placed high on the agenda when a security problem left in a software product has been publicly broken and is being exploited by attackers. Unfortunately, a patch

---

[186] This has been pointed out by on the basis of personal experience, e.g., by Preston G. Smith in From Experience 1999, p. 223 and has been verified by Baskerville et al., How Internet Software Companies Negotiate Quality, p. 55, in an interview of software developers and project managers of nine application and smaller niche software companies using Internet speed development practices. See also Carlshamre, *A Usability Perspective on Requirements Engineering*, p. 58.

[187] This is what Baskerville et al., How Internet Software Companies Negotiate Quality, p. 53, argue on the basis of interviews of software developers and project managers at nine application and smaller niche companies using Internet speed development practices.

[188] Shapiro and Varian, *Information Rules*, p. 169.

[189] Shapiro and Varian, *Information Rules*, p. 169.

is essentially the only option used when this occurs. According to John Viega and Gary McGraw[190], the problems of this pervasive 'penetrate-and-patch' approach are, among other things, that not all problems are reported to the developers that make the patches, that patches often introduce new problems because they are also rushed out as a result of market pressures, that often only the symptom of the problem gets fixed leaving the cause unaddressed, and that patches often go unapplied or are otherwise ineffective.

But it is no surprise that this penetrate-and-patch approach still is so pervasive. Not only does it help in getting products onto the market more quickly by skipping initial security considerations that seem to slow down development, diminish functionality, and give revenue only in the long run, but patches are also cheap to distribute. Distributing a patch for a piece of software is a lot cheaper than for traditional commodities, due to the immaterial nature of the good and consequently the virtually zero marginal cost of transmission. Developers can make patches available to the web or e-mail them out to customers and the cost of installing the patch falls on to the customers[191].

A fast cycle time together with a release orientation is something that is impossible to achieve in a serial process. *Parallel development* is used widely in release-oriented development: traditional serial phases are split and assigned to separate groups of developers, which then

---

[190] Viega and McGraw, *Building Secure Software*, p. 16.

[191] It has to be noted that even though the distribution of a patch is cheap, the creation of a bug fix is not. This is what the research on the ROI of security investments hinted. The development and testing of patches is expensive if it is done for even nearly all of the different configurations and environments of the customers. There is a strong customer reliance on the vendor to test the patches in many environments before their release especially if the patch is streamed (automatic update). However, since most bugs are harmless and go unnoticed this cost is relatively low compared to removing majority of bugs before release.

perform them simultaneously[192]. Quality assurance and testing are also done in parallel with other development phases; if a rapidly approaching release date forces the company to shorten development phases, quality assurance and testing also get short-circuited[193].

## 2.3 Remarks on maintenance and testing

The product focus in software development accounts for how *software maintenance* changes in the software products market. In traditional tailored software development, there usually is just one release and the fixes, additional features and other evaluations are provided as part of maintenance. In market-driven development this work is done by making new releases of the same product[194]. Even though the COTS software vendors separate corrective maintenance – including patches and workarounds which are often provided to licensees at no cost beyond a subscription fee – from other forms of software maintenance, the changes needed to smooth out poorly done but operable software functions become the basis of new releases for which vendors charge additional, often highly profitable licensing fees. In other words, maintenance takes the form of versioning or supporting services, for which the customer has to pay separately. Most of what was once maintenance in traditional software development now forms the basis of a product's next release and thus

---

[192] About parallel development in market-driven process in general, see Blackburn et al., Improving Speed and Productivity of Software Development, p. 878, and in Internet software development especial Cusumano and Yoffie, *Competing on Internet Time*, p. 14, and Baskerville et al., How Internet Software Companies Negotiate Quality, p. 52.

[193] Baskerville et al., How Internet Software Companies Negotiate Quality, p. 54.

[194] See, e.g., Natt och Dag, *Elicitation and Management of User Requirements in Market-Driven Software Development*, p. 18; Carlshamre, *A Usability Perspective on Requirements Engineering*, p. 59.

serves to generate additional revenue for the vendor over a number of years[195].

The combination of release-oriented development and patching and, especially, the use of new releases as an important form of maintenance lead to a reliance on customer feedback as a significant, or even primary, quality assurance mechanism in market-driven development. This is reasonable economic behaviour as one of the most appreciated and loud advocate of information security Bruce Schneier so effectively puts it in his famous quote "…90% to 95% of all bugs are harmless. They're never discovered by users, and they don't affect performance. It's much cheaper to release buggy software and fix the 5% to 10% of bugs people find and complain about."[196]

This has implications for quality and security. Press coverage is not guaranteed to be accurate and may not convey the implications of the problem being reported. The problems that concern only a smaller user community do not get fixed. Feedback from customers and the press, by its very nature, occurs only after a product has been distributed. Reliance on market forces to select what gets tested and what gets fixed is haphazard at best and is surely not equivalent to performing a methodical search for vulnerabilities prior to distribution.[197]

The development goal of achieving software that is 'good enough' – not perfect (flawless) – for the specific situation and customer needs also applies to *testing*. A lot of bugs are detected during testing procedures, but not all errors can be found. As Cem Kaner (among others) explains it, it is impossible to fully test a program: the testing procedure can only show the presence of errors in the program; it cannot show the absence of errors[198]. Additionally, testing cannot

---

[195] Sawyer, A market-based perspective on information systems development, p. 101.

[196] Schneier, Computer Security: Will We Ever Learn?, title "No one is paying attention because no one has to", paragraph 5. Similarly in Harju, *Kustannustehokas ohjelmiston luotettavuuden suunnittelu ja arviointi*, p. 89.

[197] Schneider, *Trust in Cyberspace*, p. 89-90.

[198] Kaner, The Impossibility of Complete Testing, p. 1.

show that the software has certain qualities. Despite these limitations, testing is widely used in practice to create confidence in the quality of software[199]. This is why software is shipped with bugs even after the verification and validation stage. However, the bugs that remain do not prevent software from being 'good enough'.

The verification and validation stage can be one of the more time consuming, expensive and challenging phases of the software life cycle. It has been estimated that about 50% of the development costs of a software product are caused by testing and debugging[200]. In market-driven development (a very competitive market), the design team is often under tremendous pressure to complete this phase. Market pressures contribute to reducing the time spent on testing before releasing software to users. More sophisticated testing and debugging procedures would prolong introduction of a new product (to the market) as well as add costs and thus decrease the probability of commercial success. In sum, software is being released and implemented without adequate testing[201].

There are special problems in testing for security. The testing procedures must be changed to focus on security issues (e.g. testing for unexpected input, probing a system like an attacker or otherwise looking for exploitable weaknesses) in order to find the particular vulnerabilities as the practitioner of security tend to point out[202]. Functional testing (treating the component as a black-box and testing the interfaces of the components) does not find security flaws. Unlike almost all other design criteria, security is independent of functionality. Functional testing is good at finding random flaws that, when they happen, will cause the computer program to behave oddly. Security flaws have much less spectacular effects; they are usually invisible

---

[199] This has been emphasised by Mary Jean Harrold in Testing: A Roadmap, p. 63, while mapping the role of testing in the future of software engineering.

[200] Harrold, Testing: A Roadmap, p. 63; Harju and Koskela, *Kustannustehokas ohjelmiston luotettavuuden suunnittelu ja arviointi*, p. 10.

[201] Similarly in Pipkin, *Information Security*, p. 75.

[202] See, e.g., Pipkin, *Information Security*, p. 74; Viega and McGraw, *Building Secure Software*, p. 39.

unless they fall into the wrong hands. Security testing is not about randomly using the software and seeing if it works, but deliberately searching for problems that compromise security[203]. The costs of testing, together with the time needed, increase when security is concerned, which is why software rarely end up being 'good enough' in terms of security even after testing.

The method used in market-driven development – enlisting the user community to help in finding errors by making early releases (beta versions) available to interested users and by freely distributing incremental updates (e.g., patches) to the software – does not enhance security, since no amount of beta testing will uncover all security flaws[204]. This is mainly due to the need for sophisticated security-specific testing to find vulnerabilities in the first place. Knowledge of the testing methods and skills to conduct them are not typically

---

[203] This is the argument is repeated at least by security practitioners such as Bruce Schneier (*Secrets & Lies*, p. 335-336), Donald Pipkin (*Information Security*, p. 70) and John Viega and Gary McGraw (*Building Secure Software*, p. 42).

[204] In market driven development the trend has for long been to enlist the user community to help in finding errors by making early releases (beta versions) available to interested users and by freely distributing incremental updates (e.g., patches) to the software. Beta testing has traditionally been part of the pre-release testing (before release to the wider customer base and not just to the interested beta-testers). Recently, however, software has increasingly been released to be public as a 'final' version, with an implicit assumption that the end-user, and not just the willing beta-tester, will act as the ultimate 'beta' tester. Rather than implementing a full quality assurance program, the vendor relies extensively on users to report vulnerabilities.
Proprietary software developers, as contrasted to open source software (OSS) development projects, especially those with COTS business model and market driven development practices of interest in this study, are increasingly turning to their customers for help in the debugging task (to remove bugs from the software). Proprietary software developers understand just as well as do open source software developers the value and potential of users in testing. The principle of release early and often makes sense in both settings. As noted earlier, proprietary software has been seen as going towards a plausible promise option in terms of the business model and thus moving closer to the OSS development method. This means that when the product is first made available to users, it is not finalized in terms of functionality or quality. The product is gradually improved in terms of quality and functionality in the subsequent releases partly due to the feedback from the public.

widely known in the broader user community. Of the parties searching for vulnerabilities, hackers and professional tiger teams may have the skills and motivation (at least to some point), but are in no way able to do it quickly and efficiently for every software product[205]. As pointed out by the Computer Science and Telecommunications Board in 2002, software vendors should "[s]trengthen software development processes and conduct more rigorous testing of software and systems for security flaws, doing so before releasing products rather than use customers as implicit beta testers to shake out security flaws"[206]. This has also been emphasised at the European policy level by the Economic and Social Committee in its opinion on the Commission Communication on Network and Information Security COM(2001)298 final[207].

A further problem in using customers (either end-users or software developers using components) as testers is that they typically do not have access to the component's source code and to the specific documentation of the production process, especially where COTS is concerned[208]. COTS vendors seeking to protect their intellectual property usually sell components as binaries, without source code

---

[205] John Viega and Gary McGraw raise the problems of red teaming for security in *Building Secure Software*, p. 42-43. Unfortunately, this method is still sometimes praised as an efficient way of discovering vulnerabilities. For example, in noting correctly that faster and often less sophisticated testing procedures allow for a shorter time-to-market, thus leading to a competitive advantage, the Finnish Ministry of Transport and Communications in a report concerning the need for national information security strategy in 2001 also stated that end-users are able to find even information security vulnerabilities quite fast (MINTC, Kansallisen tietoturvastrategian tarve Suomessa [Does Finland need a national information security strategy?], p 7).

[206] Computer Science and Telecommunications Board (CSTB), *Cyber-security Today and Tomorrow: Pay Now or Pay Later*, p. 14. CSTB is a division of the U.S. National Research Council.

[207] Opinion of the Economic and Social Committee on the Commission Communication on Network and Information Security: Proposal for a European Policy Approach (COM(2001)298 final), Official Journal C 048, 21.02.2002, p. 33-41, paragraph 3.2.1.3.13.

[208] Harju, *Kustannustehokas ohjelmiston luotettavuuden suunnittelu ja arviointi*, p. 51.

or design documentation[209]. The lack of availability of component source code limits the testing that the component user can perform (white-box techniques in evaluation of components is not possible)[210].

Even though some traditional security analysis is made impossible for the component user or other customer by the absence of source code, there are ways for the user to verify and determine the quality and security of COTS components that do not require extensive disclosure of the source code or the accommodated design documentation. There are approaches that treat the component as a black box, and employ extensive testing to ensure that the system functions as desired; no additional effort or disclosure of intellectual property rights (IPR) are required from the COTS vendor[211]. Grey-box

---

[209] The legitimate reason for this is that developers want to keep source code forms of their products and other human-readable documentation as trade secrets as, e.g., Pamela Samuelson and Suzanne Scotchmer point out from the U.S. perspective in *The Law & Economics of Reverse Engineering*, p. 1608, similarly to Alfred Meijboom's notion from the European perspective in Legal Rights to Source Code, p. 107. The procurement policies of the customers of safety-critical systems (utilities, government, etc.) have traditionally required software vendors to disclose enough details to evaluate their processes and products for safety. However, these policies are not compatible with current component vendors, who are faced with the risk of IPR loss as pointed out by Devanbu and Stubblebine in Software Engineering for Security, p. 233. Not only are software components delivered in "black boxes" as executable objects without source code or design documentation, but usually also the de-compilation back to source code is forbidden in licenses (note that this is forbidden in copyright law, but not in trade secret law). Often source code can be licensed, but the cost may make the practice prohibitive as Jeffrey Voas point out in 1998, Certifying Off-the-Shelf Software Components, p. 53. The open source and free software movements offer the source code to the users and also commercial vendors are, at least to some degree, starting to show their source code to trusted partners (e.g. governments) also for security purposes.

[210] It needs to be pointed out that code evaluation is a necessary but not a sufficient means for assessing security as emphasised by John Viega and Gary McGraw in *Building Secure Software*, p. 115. Security related vulnerabilities can be found even without a look at any code (source of binary) – in worst cases symptoms of a security problem are noticed during the course of normal use (Viega and McGraw, *Building Secure Software*, p. 70-73).

[211] Such methods have been reported by Jeffrey Voas already in 1998, Certifying Off-the-Shelf Software Components, p. 53-59.

verification systems use interactive cryptographic techniques or rely on tamper-resistant hardware to help the vendor to provide evidence of the quality and security of the component (disclosure of enough details of the verification practice to convince a sceptical component user) without disclosing too much information that could endanger its IPR[212]. There are also different sets of criteria on which components are evaluated and verified (e.g. ITSEC, TCSEC a.k.a. Orange Book , Common Criteria).

Even though the additional testing effort required by black-box approaches contributes towards the overall quality of the component user's entire system, their use is limited because the additional testing is likely to be time-consuming and expensive[213]. An additional limitation on the use of black-box approaches is that they do not reveal unknown, malicious functionality[214]. Grey-box approaches have only very recently appeared and need a lot of additional research. However, with additional research into the ways in which component users can test systems efficient techniques and tools are likely to emerge that will help such users test their applications more effectively[215].

---

[212] Devanbu and Stubblebine, Software Engineering for Security, p. 234.

[213] Reifer, Boehm and Gangadharan in Estimating the Cost of Security for COTS Software, p. 180, note that current models for predicting the effort involved in integrating COTS software products into applications do not include security as a cost driver. While providing means for estimating the costs of security for COTS software they also (idem. p. 183-184) estimate on the basis of their analyses the percentual increases both of the effort and the duration of the assessment activity (process by which COTS components are selected for use; 12-20 percent to effort and 5-10 percent to duration), to tailoring (activities undertaken to prepare the selected COTS packages for use; 8-18 percent to effort and 5-10 percent to duration) and to glue code development (development and testing of the connector software, which integrates the COTS components into the larger application; 0-75 percent to effort and 0-33 percent to duration).

[214] Voas, Certifying Off-the-Shelf Software Components, p. 55.

[215] Whether these problems can be solved by technological means, or are they more about simple economic decisions made by the developers of both the components and their users, thus possibly requiring some sort of regulatory

The use of customer feedback in place of other quality control mechanisms does allow a software producer to externalise costs associated with product testing. Customers, in turn, have to invest time and money in finding and possibly reporting errors, in installing patches, and they also have to suffer from the costs of failures. The tactic of using customers as serious (perhaps involuntary) testers is, at best, a dubious one from the point of view of security[216]. It is surely not equivalent to performing a methodical search for vulnerabilities prior to distribution[217].

## 2.4 Appeal to developers and security

When competing to dominate the network market, i.e. to achieve the critical mass needed to take it over, firms have to appeal strongly to the developers of the next-generation 'killer-aps' and vendors of complementary goods and services[218]. Good developers are needed to create the products that attract customers in the short timeframe required and vendors of complementary goods help to build up the critical mass of users for the product and its applications. Also the appeal to early adopters has to be strong since they are the critical mass – if they are pleased. One can worry about the rest of the end-users later.

This helps to explain why the security features in commercial software, if they are even implemented to begin with, are made easy

---

intervention, is a matter of dispute.

[216] This argument is made in one of the first ever textbooks on secure software development, Viega and McGraw, *Building Secure Software*, p. 17.

[217] Schneider 1999, p. 89-90; Viega and McGraw, *Building Secure Software*, p. 17; CSTB 2002, p. 13

[218] This argument has been made in the relatively recently emerged economics and information security -discussion, e.g., by Ross Anderson, *Cryptology and Competition Policy-Issues with 'Trusted Computing'*, p. 14, and in Anderson, *Why Information Security is Hard*, p. 3. The argument has, however, been made already by Andrew Odlyzko already in 1998, Smart and stupid networks, p. 38-46.

to bypass. Stronger security that is not easy to bypass may diminish functionality and require more work from and less opportunities (increased costs and diminished revenues) for the developers of the next generation 'killer-aps' and developers of complementary goods and services. Since early adopters typically are persons with high technology skills that want to try out and test new things, reduced functionality together with security that is not easily bypassed, might lower the appeal of the product among this conceivable critical mass. Due to the relative unimportance of end users (they just want to get a few crucial tasks done) and the crucial role of the developers that create the tools that attract the early adopters, the cost of insecurity (caused, e.g., by vulnerabilities being abused and when upgrading the system) and other support costs needed to operate with the software (from implementation, maintenance, testing and dealing with the failures) is dumped on to the end-users. These costs have long been hidden.[219]

In an attempt to appeal to the early adopters and thus to exploit the positive feedback, vendors may add features to attract this small number of users with special needs even if those features will be unused by other users. This can result in lower security due to an increase in the number of vulnerabilities, the usual lack of testing in these unused parts, and the security problems incurred when integrating these additional features into the system.

Because consumers typically possess imperfect information about the product's security, vendors can get away with both of these practices. They can develop easily by-passable security features and add features with increased number of vulnerabilities.

## 2.5 Security and lock-in

With network effects the costs of coordinating a large group of individuals to switch to a competing product can be extremely large;

---

[219] Anderson, *Cryptology and Competition Policy-Issues with 'Trusted Computing'*, p. 14; Anderson, *Why Information Security is Hard*, p. 3.

the effects contribute to customer *lock-in* (i.e., the costs of switching from a product to another are so large that switching suppliers is virtually unthinkable)[220]. And because the building of network size by a competitor with an incompatible product requires overcoming the collective switching costs – the combined switching costs of all users – customer lock-in becomes the norm in the information economy[221]. If an installed base of users is established before the competition arrives on the scene, achievement of the scale economics necessary to compete can be made difficult for later entrants[222]. Thus, first-mover advantage is not only powerful but it can also be long lasting in lock-in markets.

One implication of the first-mover advantage together with lock-in in software markets is that the market may settle on a good with a lower *social* valuation. Once the market tips toward a single standard, it may remain on that standard and its successors for a long time even though an objectively 'better' standard is available. Even though all users would be better off with the new standard, those benefits do not accrue to the present users, who would have to pay substantial switching costs. New purchasers also may opt for the established standard because of the immediate benefit that the established network offers; they do not take account of the benefit that purchasing the new product would confer on later purchasers. Even if they anticipate that the new product will be widely adopted, the benefits of that adoption to new purchasers may be realised so far in the future that they are substantially discounted.

---

[220] Network effects are a common source of switching costs: when a product has become ubiquitous it is very costly to switch to something new (Shapiro and Varian, *Information Rules*, p. 47). Internet distribution of new applications and standards reduce some of the network effects for software by reducing switching costs. Variety can also be supported more easily if an entire system can be offered on demand. However, the Internet does not eliminate network effects in software; interoperability is still a big issue on the supply side and there still is a strong need for standardisation. (Shapiro and Varian, *Information Rules*, p. 189-190)

[221] Shapiro and Varian, *Information Rules*, p. 110.

[222] Shapiro and Varian, *Information Rules*, p. 168.

In the present context, this theoretical possibility means that the market may get locked-in into a product that is insecure (has security related vulnerabilities or lacks security features) even though a more secure product is available in the marketplace[223]. This is because users do not, in network markets, choose software purely on the basis of its features and security. Network effects are also significant, because the utility of software product increases with the number of users and users of compatible products. A typical user may need a certain set of features and beyond that be concerned with standardisation and interoperability. When a market has tipped in favour of one product, users who, in autarchy, would be willing to sacrifice functionality for security may choose a less secure product because of the benefits of interoperability. Because of network effects, users who would otherwise prefer increased security to increased functionality might then choose less secure but more widely used programs. There is no expectation in favour of markets getting locked into an inferior (less secure) product, but it is a possibility that has to be taken seriously: when it actually occurs, it has non-trivial consequences, and there exists at least one other feasible state of

---

[223] The result may be a near monopoly situation, which in turn has been considered to cause new set of problems also for information security. Since the concentration in this study is on competitive markets, the monopoly situation is not taken further. For the occasionally heating debate surrounding Microsoft and its software products see, e.g., the report written in the name of the Computer & Communications Industry Association (CCIA, http://www.ccianet.org [8.3.2006]), a lobby organisation for U.S. based companies in the computer, Internet, information technology, and telecommunications industries, in 2003 by Daniel Geer, Becky Bace, Peter Gutmann, Perry Metzger, Charles P. Pfleeger, John S. Quarterman, and Bruce Schneier, *Cyber*In*Security: The Cost of Monopoly. How the Dominance of Microsoft's Products Poses a Risk to Security*, available at http://www.ccianet.org/filings/cybersecurity/cyberinsecurity.pdf [8.3.2006] and the discussion in the November-December 2003 issue of the IEEE Security & Privacy written by Daniel E. Geer Jr., Dave Aucsmith and James Whittaker (2003) Monoculture, *IEEE Security & Privacy*, 1(6): 14-19.

affairs that might be preferable as tentatively noted by Tom Lookabaugh and Douglas Sicker in 2003.[224]

In the case of market-driven development, lock-in into a vulnerable product is likely. As discussed above, with first-to-market as an essential feature in achieving a market share needed for bearing the competition, security tends to become an afterthought. However, the winner ought to make its product better in quality and security over time. Yet this might take a long time or not happen at all, because the code of the first versions has emphasised functionality over security and a complete rewrite of it is not practical given the evolutionary model of development: development consists of modifying previous versions and, over the years, these become so complex that they simply could not be developed (or redeveloped) from scratch[225].

Thus, the design space of release-oriented development is highly influenced by the state of the code after the previous release(s). Because adding security later in the development phase is difficult and expensive, and when previous release(s) has (have) emphasized functionality over security, increasing the security of it in a new release requires significant resources that will not then be available for enhancing functionality. But, since functionality still has to be improved in order to make the new release attractive (due to heavy competition), the resources are easily taken from features that are considered less useful (e.g. quality and security). This means that initially insecure code does not necessarily improve with new releases.

---

[224] Lookabaugh and Sicker, *Security and Lock-In*, p. 1. According to theoretical models, there is no inevitable tendency of markets to lock-in on inferior products. Even though the theoretical possibility of such lock-in is not contested, the empirical evidence and the practical importance is (Idem., p. 4).

[225] Robert Brady, Ross Anderson and Robin Ball make this argument in *Murphy's law, the fitness of evolving species, and the limits of software reliability*, p. 5, while developing a reliability growth model for software. Cf. Baskerville et al., How Internet Software Companies Negotiate Quality, p. 53, who report the practice of gradually rewriting the code of the first release in order to meet the contemporary quality requirements.

The positive aspect of lock-in is that software and content producers in general have an increased interest in security mechanisms for their value in locking customers into their products and systems[226]. This can be achieved by using information security protocols to set technical compatibility requirements that must be met by connected applications. A supplier wishing to sell a system component will either need to be compatible with the necessary security protocols (by licensing or reverse engineering), or must provide sufficient added value to motivate a customer to replace all other system components that require those security protocols, which may result in prohibitive switching costs[227].

This also has consequences in the legal field. The strong and effective lobbyists of the rights holders and software vendors thus have incentive to try to have laws enacted that protect their commercial interests (from illegal copying) and, at the same time, silently enhance customer lock-in; all one has to do is look at the anti-circumvention rules in the EC Copyright and Conditional Access Directives and the US Digital Millennium Copyright Act[228]. When security is brought up, the rights holders worry about the safety of their respective intellectual property assets and possibilities to keep their market share but do not care much about the security of the

---

[226] For the argument see, e.g., Lookabaugh and Sicker, *Security and Lock-In*, p. 7; Anderson, *Cryptology and Competition Policy-Issues with 'Trusted Computing'*, p. 14.

[227] Note that reverse engineering security systems (conditional access devices) has been made more difficult by the anti-circumvention rules (e.g. in EC Copyright and Conditional Access Directives, and US Digital Millenium Copyright Act). It is practically impossible to reverse engineer a technical protection measure without circumventing it. Because reverse engineering a technical protection measure usually requires a tool to perform such activities these provisions indirectly restrict reverse engineering by outlawing the making of (or other 'have something to do with') circumvention technologies.

[228] For an excellent overview of the regulatory background, see the analysis of Peter Drahos and John Braithwaite in *Information Feudalism*, p. 184-186, drawing on data from interviews of key informants of the rise of the TRIPS agreement (Trade-Related Aspects of Intellectual Propery Rights).

underlying information infrastructure[229]. This attitude poses a critical problem for security in a networked world: laws that protect possibly insecure software are going to protect network insecurity at the same time[230].

## 2.6 Failure of private motivation?[231]

Secure software, both in terms of the quality aspect (avoidance of vulnerabilities) and security features, is a commodity bought and sold on the market and part of the contractual agreements between parties. The implicit assumption usually made is that the price mechanism will balance the costs of providing security with the specific need for security. Certain users will request high security whilst others will be satisfied with a lower level – although the State may provide for a minimum level of security in certain specific contexts[232]. The preferences of users would be reflected in the price they are willing to pay for both the quality aspect of security and the security features. But this assumption seems imperfect when looking at the many security risks that remain unsolved or the slow market entrance for solutions as a result of certain market imperfections[233].

---

[229] Both of these aspects, using security technologies to protect IPRs and to secure the infrastructure, are about information security and can use similar technologies and procedures. However, the security measures serve conflicting interests. This makes information security research difficult; there are dual uses in many senses and the security measures can be used for various purposes.

[230] This argument has been made, e.g., by Robert Gehring, *Software Patents" — IT–Security at Stake?* 2001, p. 2-3.

[231] Please, bear in mind that the analysis concentrates on the abstract situation present at the turn of the millennium where regulatory pressure has not initiated any substantial changes.

[232] However, in case of secure software this is rare.

[233] This has also been recognised at the policy level in the EU (COM(2001)298 final, p. 4 and 13-14). See also the opinion of the Economic and Social Committee on the Communication, Official Journal C 048 , 21,02.2002, p. 33-41. The market for information security has long been seen as dysfunctional.

As have been argued above, the industry has the means to develop more secure software products (both in terms of quality and security) and they at least should have a very powerful incentive to try to avoid the defects as early as possible (cost, productivity and time-to-market benefits). With the wide acknowledgement of the importance of developing security into systems from the beginning, its ability to enhance both productivity and speed to market and even the cost savings possible related to it (it is cheaper in the long run to fix security vulnerabilities sooner rather than later in the development life-cycle), why there still is not enough of an incentive for application developers to stop releasing insecure code?

The short expedition into the world of information economics and the market-driven development showed some limitations to the incentives of software producers to improve the quality and security of their products. It is not that software houses are incapable of producing systems with fewer vulnerabilities, it is just that they are not sufficiently motivated to do so. Applying the methods to improve software quality and security can be resource intensive and will not be done in market-driven development without sufficient incentive. Only as much resources are allocated to developing more secure and better quality systems as can be justified on business grounds. In the presence of network effects this might not be enough for the society.

This holds also a positive implicit argument basing its claim on welfare economics. Changes in the private motivation to provide secure enough software and information systems can lead to a more favourable equilibrium. The sub-optimal equilibrium, in which a vendor can be successful without improving security, continues as long as customers accept the situation, the primary demand for functionality only (the main purpose of a computing or a communication device or a system) continues to grow and fuel

---

The System Security Study Committee of the Computer Science and Telecommunications Board of U.S. National Research Council pointed this out already in 1991 (Computer Science and Telecommunications Board, *Computers at Risk*, Chapter 6 "Why the Security Market Has Not Worked Well", p. 143-178).

demand for features[234], and developers do not consider the current release's possible influence on options in future releases. Network effects do not prevent this unless the participants in the market fail to internalise them – and even then the incurred 'network externalities' are not likely to have serious consequences on security[235]. Despite the current economic disincentives for the development of secure software, the *private motivation* to provide software that is secure enough can raise the level of security. The market actors themselves are enough to improve security.

The economic consideration also imply that no amount of technology or methodological improvement (even acknowledging the need for and importance of further work on them) is able to give satisfactory solutions to security; the presence of network effects as such is enough to pose serious economic disincentives for the use of the technological tools and development methods. However, this is a premature conclusion. Both technology and the development methods can be harnessed for regulatory purposes (i.e., used intentionally to alter the behaviour of the developers)[236]. They can

---

[234] Schneider, *Trust in Cyberspace*, p. 188.

[235] Recall from above the differentiation of network effects and network externalities. Network effects should not properly be called network externalities unless the participants in the market *fail* to internalize these effects. Thus network effects are not market failures (failures to provide socially optimal solutions) as such.

[236] Technology and procedural rules can be approached to in many ways. The traditional approach is technological development in the shadow of market processes; competitive market forces develop new solutions in order to be able to better meet the demands of the customers. Technology is developed within the market incentives and the development methods support the market needs. This is the approach when making the implicit argument that technological means and development methods are not sufficient to provide secure software without proper incentives. Under this conception, technology provides a solution only to certain degree… But technology can also be used in a regulative manner. It can be put it in place in order to intentionally attempt to alter the behaviour of others. In this sense the technological measures and development methods are instruments in the regulation of software development. They can be altered, if the regulator so desires, to consider the incentives or even provide proper incentives. However, this does not happen

be altered, if there is desire, to change the effect of the economic disincentives. Fortunately, the seriousness of the economic disincentives has been recognised and demands for their correction are emerging.

Viewing secure development in a longer term (over the current release) would make security and quality aspects seem more desirable for the producer. Each individual release should not only address the immediate requirements for the current design, but also concern the release's possible influences for the options in future releases. For example, in order to obtain increased security in a future release, it might be necessary to build the current release with more security and less functionality than what the immediate requirements calls for. Unfortunately, this is not the case in market-driven development due to time-to-market being so essential in conquering market share.

The most promising change is that issue of software vulnerability and lack of security features is getting wider and wider attention. The hope is that customers no longer tolerate the poor quality and lack of security, and learn to demand also other than just functional requirements; i.e., they really start to demand for quality and security. When customers are in a position to switch products and services in a competitive environment, then market forces will eventually eliminate nonperforming suppliers.

Unfortunately, there are serious hindrances. In many cases there are no secure alternatives to which user could switch to. The vulnerability of software is a common problem for the whole software industry. It is not dependent on certain vendors. And even if more secure alternatives enter the market, the lock-in effect still hinders their wider adoption.

Also customer and end-user expectations still seem to be fulfilled due to the acceptance of a reasonable degree of insecurity and unreliability or operational difficulty as a trade-off for innovation. As the near term history of market-driven software development

---

without the intention to do so and the possibility to influence the behaviour of others.

explains, COTS software development practices in the personal computer (PC) era arose in a technical and economic environment that was very accepting of errors and malfunctions[237]. The acceptance of errors and malfunctions still exists to some degree. Customer and end-user expectations seem to be fulfilled due to the acceptance of a reasonable degree of insecurity and unreliability or operational difficulty as a trade-off for innovation[238]. Customers have demanded functionality, not security. The primary demand for functionality only (the main purpose of a computing or a communication device or a system) has continued and fuels demand for features[239].

Yet, this tolerance of poor quality and lack of security will diminish with time (perception of product as less innovative) and competition

---

[237] PC operating systems and applications used to ran on isolated desktops; the consequences of failure were limited to destruction of perhaps valuable but certainly not life-critical data and failures had no way of propagating to other machines. This climate was amplified by economic conditions of the early PC era. Software was purchased separately rather than being bundled with a leased computer, as in the mainframe era. Consequently, there was less financial leverage for dissatisfied customers to affect vendor, and therefore developer, attitudes. A customer's financial leverage was limited to consuming vendor resources in calls to telephone help-lines, which could be ignored by inept or uncaring vendors, and refusing to purchase other software or the next revision of the malfunctioning product from that vendor. The latter option is reduced by the diminishing diversity of the marketplace, the need to exchange data with other users, and the investment the customer may have in data that can be processed only by the product in question. (Schneider, *Trust in Cyberspace*, p. 88)

[238] Greater risks are usually tolerated where highly socially useful products are provided. Clearly, software has proven enormous utility despite the costliness of attacks and this is why the related intrinsic risks are tolerated. However, the current state of security can be changed and greater security can be achieved as has been argued above. The equilibrium, if such even exists, is not optimal. As Jennifer Chandler notes in *Improving Software Security*, p. 20-21, the relevant inquiry is whether the cost of improving security is more or less than the value of the improved security. Even though this line of argumentation is not taken further in this study, the basic argument is that at least in the case of most common vulnerabilities the improvement of security is cost-justified from the societal point of view.

[239] The situation has not changed all that much from 1999 when Fred B. Schneider made the argument in an influential National Research Council report *Trust in Cyberspace*, p. 188.

(competitors may introduce new products and releases with new features, and users will very quickly lower their tolerance if a competing product does better)[240]. This is what we are currently experiencing; customers have started to put pressure also on proprietary software vendors on security issues and they have reacted accordingly[241]. As more and more aspects of society become dependent on computing, customers have increasingly started to demand dependably secure software[242].

Vendors naturally are very keen to provide what the potential customers' desire with respect to security and quality of their products because companies that fail to deliver on the customer requirements may soon find themselves without any customers. Commercial software companies therefore seem to have, at the first sight, a very strong incentive to ensure the level of security their customers demand, and they must be willing to invest whatever resources are necessary to do so[243].

---

[240] The dominant players can delay the effect of competition in particular, because high switching costs cause customer lock-in into current products.

[241] The trend of demanding security of software has been reported in the U.S. already in 2002, e.g., by Dennis Fisher in an eWeek article Contracts Getting Tough on Security available at http://www.eweek.com/article2/0,1895,1658531,00.asp [29.3.2006]. Whether this, in the final end, is a natural development in competitive markets or due to increasing threats of regulation pressuring change is part of the analysis conducted below.

[242] However, vendors have also found that, although a customer might claim that security and quality are important in abstract, when it comes time to spend money, functionality and performance expenditures often take precedence. This argument has been made, e.g., by Fred B. Schneider in *Trust in Cyberspace*, p. 198, and has been repeated by Alessandro Acquisti and Jens Grossklags in *Losses, Gains, and Hyperbolic Discounting.*

[243] Note that not all of the users have to start to demand security and actually be ready to pay for it. As long as vendors cannot differentiate informed customers from uninformed ones, they have an incentive to produce quality. That is, if the number of informed customers is big enough. Thus, it suffices that the central users (the critical mass or the vendors of complementary goods and services) or another big and powerful enough a group starts to demand security and quality. Of this argument in general see the analysis of Ejan Mackaay, *Economics of Information and Law*, p. 149-150, in an effort to introduce

Basically, private markets themselves can resolve many of the problems discussed above - there is no direct need for intervention by the government or by another regulator. The laws of market can solve many of these problems. But is that private motivation, even if it gets stronger, sufficient to provide an optimal amount of protection for *society* as a whole? Are the market actors themselves sufficient to provide secure software for the needs of society and not just customers?

This is where we turn to welfare economics[244] and other disciplines that explain why problems might occur. This line of microeconomic theory suggests other reasons, in addition to the effects of the network economic environment where the market-driven development works, why the industry might fail to provide secure enough software from the societal point of view: *externalities* (spill-over effects) affecting third parties in ways not reflected in the price set by producers or in cost considered by the buyers, and *inadequate or asymmetrical information* affecting the relationship between suppliers and buyers. Co-ordination problems have consequences for both of these market failures (e.g. though desired outcomes can in principle be achieved by private transactions at the presence of externalities and asymmetric information, the costs of co-ordination are so high that it is cheaper for the law to prescribe conduct).

The existence of market inefficiencies such as externalities and imperfect information suggests that the present level of security is inefficient. These market failure considerations suggest that even if the private motivation to enhance security amends, it might not suffice to provide an optimal amount of security for society as a whole.

The following market failure considerations increase the understanding of the behaviour of the software vendors. They show

---

the issues of the costliness of information to the economic analysis of the law in 1982.

[244] Welfare economics as depicted, e.g., by Robert Cooter and Thomas Ulen in *Law and Economics*, p. 39, is the part of microeconomics theory that explores how the decisions of many individuals and firms interact to affect the well-being of individuals.

reasons why, despite the increasing demand for quality and security, sufficient improvement might not happen without intervention from the outside. They also provide a framework for understanding the effects of interventions into the market for information security and secure software especially. The considerations of the regulatory capability of specific regulatory tools build partially on this understanding.

These short considerations further justify the step into the regulatory questions; they justify the importance of looking into regulation[245]. They do not justify the need for regulation as such, especially not alone and without further considerations, for example, of the failings of the regulatory solutions themselves, of the efficiency of the tools, of the distributional and other goals beyond efficiency[246], or of the real motives for regulating beside these technical justifications stemming from market failures[247]. But they do partly explain the current situation and set the basis for looking what kind of instruments (influence mechanisms and incentives) regulation can provide for the development of secure software. They further justify why incentives from the outside are relevant and why they are needed. The analysis also makes explicit that the reasons for regulating secure software development is a combination of rationales even under pure market failure analysis.

---

[245] Wolfgang Kilian in one of the first conferences ever considering the role of law in securing computer networks already justified the need to raise the question of the role of legal provisions with market failure considerations. Kilian was in Data Security in Computer Networks and Legal Problems, p. 16, concerned that transaction costs and efficiency considerations will take precedence over technical security measures.

[246] These are presented tentatively by David Weimer and Aidan Vining in their textbook, *Policy Analysis*, p. 134-196.

[247] As Robert Baldwin and Martin Cave explain in their textbook on regulation, *Understanding Regulation*, p. 9, motives for regulating need to be distinguished from technical justifications for regulating. The latter are the market failure explanations given by a regulator that is assumed to be acting in pursuit of the public interest. The former are the actual motives behind these technical justifications that may stem for example from the interests of the regulated industry or be influenced by the economically powerful.

Please, bear in mind that where the above analysis of the implications of network effects on secure software development implicitly supported the argument that technological or procedural solutions *as such* are not sufficient to provide security due to the disincentives in their use, it was noted (on the basis of the wider understanding of regulation) that technology and development methods can also be used 'regulatively'.

Similarly in here the argument made with the market failure considerations is not to imply that governmental intervention is needed or to recommend what the public policy ought to be. The decentring thesis accepted in this study widens the regulatory field essentially. It is not only the government that can intervene in the markets; also other regulators with their capacities can alter behaviour and they can use the same instruments (excluding the use of force which is a state monopoly). Thus, the market failure arguments made below are not meant to speak in favour of governmental intervention. They speak about the need for regulatory incentives, but they can come from a variety of sources.

## 2.7 Information security as an externality

Information in general is a special commodity. On the demand side, buyers cannot determine the value of information (how much they would pay for it) until they have it, and having it removes their willingness to pay for it. There are differences also in the supply side: information is costly to produce, and yet it costs relatively little to copy and transmit. Thus, it is extremely hard for anyone who has devoted resources to the production of information to appropriate its value through the sale of that information. This is because the buyer of the information can resell it at the cost of transmission. Owing to the low cost of transmitting information, information producers have difficulty selling information for more than a fraction of its value (economics call this the problem of non-appropriability). Consumers desire to become "free riders" for information, paying no more than

the cost of transmission for the commodity (e.g. copying a computer program for free).[248]

Why is the appropriation of the value of information so difficult? An answer can be found from the theory of public goods (economics see the problem of non-appropriability similar to the public goods issue). Public good is a commodity which benefit is shared by the public as a whole, or by some group with it. It has two very closely related characteristics: non-rivalrous consumption (i.e. consumption by one person does not leave less for any other consumer) and non-excludability (i.e. the costs of excluding non-paying beneficiaries who consume the good are so high that no private profit-maximizing firm is willing to supply the good)[249].

> As Patrick McNutt explains in the Encyclopedia of Law and Economics[250], a pure public good exhibits in extreme measure the characteristics of non-rivalry in consumption (one person's consumption of it does not diminish the amount that others are able to consume), and non-excludability (no one can be excluded from enjoying it). The property of non-rivalrous consumption implies zero marginal cost to existing users in sharing the benefits of the good with an additional person. If a pure public good is privately provided, then it will quickly be provided at that zero marginal cost; but at that price it does not pay any private producer to supply it since their investment (the fixed costs of production) could never be recovered. At the same time, the non-excludability of public goods implies that the benefits of access cannot be fully appropriated by producers: if someone cannot be excluded from enjoying the benefits of a good, he or she has little incentive to pay for it but will be inclined to "free ride". There is then little incentive for a private producer to undertake the supply of such a good, which would consequently be undersupplied.

---

[248] Cooter and Ulen, *Law and Economics*, p. 109 and 126.

[249] Cooter and Ulen, *Law and Economics*, p. 42.

[250] McNutt, Public Goods and Club Goods, p. 927

Firstly, information contains ideas and person's use of an idea does not diminish its availability for others to use, so there is *non-rivalrous* consumption of information. Possession of certain information, e.g. using computer software, leaves it still equally valuable to another individual because she can have it at the same time. Secondly, excluding some people from learning about a new idea can be expensive, because the transmission of ideas is so cheap – the use of information is *non-excludable*.[251]

But this should be nothing new, not even to lawyers, since the copyright theories and the discussions on freedom of information and publicity principles are based on the public goods problem. But, does the public good nature of information create similar problems to the security of it, in other words, does information security exhibit in extreme measure the characteristics of non-rivalry in consumption and non-excludability similar to information as a commodity?

Actually there is no clear cut answer. In order to make a clear point in relation to the quality aspects of information security we must come back to the general information security question for a while, even though we have earlier limited the study to concentrate only to the development of secure software part of information security. This is because the different parts of information security require different treatment under welfare economics and without shortly considering them it is not possible to make a clear point about the development part of information security. There is however, a need to introduce the concept 'externality' into the discussion before we can continue because economic theories of public goods and externalities caused

---

[251] These considerations suggest that the unregulated market will produce sub-optimal amounts of information (undersupply), such as in inventive ideas and in creative works. And this, in turn, suggests the need for governmental intervention in the market for information. Even though the view that unregulated market will undersupply information is still dominating most policy discussions, situations can occur in which no regulation results in too much information or just the right amount as explained by Robert Cooter and Thomas Ulen in *Law and Economics*, p. 127, and Anthony Ogus in *Regulation*, p. 40. However interesting this wide discussion is, it will not be taken further in this context due to the lack of direct connection to the issue of information security.

by consumption or production of some good are closely related. Public externalities[252] typically also have these characteristics of non-rivalry and non-excludability; the external costs or benefits are not depleted when one person suffers or gains their effects.

All goods share the characterised of rivalry in consumption and excludability from supply to some degree[253]. If a good is rival, only one person can consume it at a time; if it is non-rivalrous, many people can enjoy the good without affecting the enjoyment of others. A good is excludable if the person in possession or the producer can exclude anyone from enjoying it; it is non-excludable if is impossible or too costly for the supplier to exclude those who do not pay from the benefit. For example, once property rights are defined over private goods, they are relatively cheap to enforce (e.g., the owner can exclude others from using them at low cost). With public goods, it is costly to exclude anyone from enjoying them.[254]

---

[252] Economic theory makes a distinction between public and private externalities (Cooter and Ulen, *Law and Economics*, p. 40 and 110). If the external cost or benefit affects a relatively small number of third parties, the externality is said to be a *private externality*. In such cases, it is more likely that the externality can be accounted for through private agreements. So, there typically is no market failure at the presence of private externalities since private agreement possibly with the help of private law measures is able to correct them. If the external cost or benefit affects a relatively large number of third parties, the externality is said to be a *public externality*. In such cases private bargaining is likely to be too costly, especially due co-ordination problems. I will concentrate mainly of public externalities since they raise the real problems for society.

[253] Not all goods exhibit the characteristics of non-excludability and non-rivalrous consumption to the same degree. Most public goods typically fall somewhere between the extremes of excludable/non-excludable and rivalrous/non-rivalrous, and can be called *impure* public goods as Anthony Ogus point out in an influential book developing an extensive theory of regulation based on legal scholarship and economic research, *Regulation*, p. 34. Actually majority of the real world's property lies in between purely private and purely public goods as Ugo Mattei state in *Comparative Law and Economics*, p. 52. This has implication for the corrective public policies as discussed on the following headings.

[254] Cooter and Ulen, *Law and Economics*, p. 106.

> Exchange (i.e., trading or agreeing on a bargain) inside a market is voluntary and mutually beneficial. Typically, the parties to the exchange capture all the benefits and bear all the costs. However, sometimes the benefits or the costs of an exchange (a private transaction) may spill over onto other parties than those explicitly engaged in the exchange. Because market transactions are voluntary, these spill-over effects are outside the market system of exchange and, as a result, are not considered in the determination of the market price. This is why they are named externalities[255].
> Externalities are a category of external effects that are by-products of an activity that influence the production of other goods or the welfare of other individuals. As Jules Coleman explains it externalities are inefficient external effects; social costs or benefits that result in inefficient production or non-optimal distributions of welfare[256]. In order to ensure that an efficient amount of the item is traded, there is a need to somehow internalise the externality. That is, there is a need to ensure that the external costs and benefits are considered in the determination of the transaction price[257].

*Externalities in information security in general.* Yes, information security at the level of information infrastructures (e.g. communications and electrical networks) and other networked information systems is interdependent as denominated by Howard Kunreuther and Geoffrey Heal: your security can be compromised by the failure of others to act even if you take appropriate precautions on your own. The security of system depends on the effort of many parties[258]. The lack of security in one system can cause adverse effects on others (e.g. when a virus

---

[255] Cooter and Ulen, *Law and Economics*, p. 40 and 110.

[256] Coleman, *Markets, Morals and the Law*, p. 76. The same distinction was made also about network effects and network externalities above.

[257] As Coleman notes in *Markets, Morals and the Law*, p. 76, internalisation need not, and often does not, require that the external effect itself is eliminated. Only the inefficiency in production or exchange that the externality generates is eliminated.

[258] Kunreuther and Heal, Interdependent Security, p. 231-249. Hal Varian argues similarly about system reliability in *System Reliability and Free Riding*, p. 1.

infects or a hacker breaks into one computer in a network, the whole network easily gets contaminated).

Losses from security breaches at the level of information infrastructures and other networked information systems can be dealt with only if a large number of parties coordinate to make the needed investments[259]. So, the incentive to invest in infrastructure security is affected by the security investments taken by others, because the security level an organisation can achieve is affected by the security level of others in that network. In this situation, investing in protection produces involves spillovers that result in positive externalities for the whole network. For example, when one system owner in a network (e.g. a telecommunications operator) takes additional security measures to protect her machines and networks, the overall security of telecommunications network becomes more secure. But the security of the network does not have to actually improve since the investment made by one party may involve spillovers in customer confidence (increased trust)[260].

This benefit given to others by securing one's own networks is not considered when deciding the amount of investment made to for the security measures and too little investment into security is made[261]. Because a secure system does not allow users to do any more

---

[259] This has been noted by the European Commission in its communication on network and information security (COM(2001) 298 final, p. 14). It is not required that everyone cooperates. But co-operation only works if a critical mass of players participates which is difficult to achieve as there are 'free-rider' profits to be made.

[260] When considering the customer confidence aspect, not only the information security investments made, but also sharing of security information can involve spillovers, which result in positive externalities for the industry as a whole. Enhanced customer trust in transacting with a particular firm also expands the overall market size within the industry, as the Amazon.com case makes explicit especially in the online market for books. This has been tentatively explained by Esther Gal-Or and Anindya Ghose in *The Economic Consequences of Sharing Security Information*, p. 3.

[261] In other words, investing in security seems to buy less for the firm making the investment when there is the possibility of contagion from others than in isolation as explained by Kunreuther and Heal in Interdependent Security, p. 8-

than an insecure system, system and network operators in private sector spend only as much on security as they can justify on business grounds – and this may be much less than the society needs as a whole. Further, because serious cyberattacks are rare, the payoff form security investment is uncertain. In many cases, it is society (or other users) rather than any individual firm that will capture the benefit of improved security. As a result, system and network operators tend to underinvest in security.[262]

Developing secure information and software systems can be seen as having the characteristics of a public good to some degree; maybe not in the extreme form (i.e. it is not a pure public good), but it creates positive public externalities that are not internalized and thus too little of security is provided[263]. According to the economic theory of public goods, the market fails in that it produces too little of the commodities (network security in our case) due to the lack encouragement to invest into the security of networks above the needs of that particular

---

9, andy by Kunreuther et al., Interdependent Security, p. 2. This is emphasized by the benefits given to others not being internalised.

[262] This has been explained by the Computer Science and Telecommunications Board (CSTB) report *Cyber-security Today and Tomorrow: Pay Now or Pay Later*, p. 9.

[263] As Soo Hoo et al. note in Regional Interest Group on Information Security, p. 1, information security is a part of the general public good of a secure information infrastructure, regardless of whether the information networks that provide public goods such as emergency services, defence, or basic infrastructure components are publicly or privately owned and operated. See also Camp and Wolfram, *Pricing Security*, p. 31-39, where information security is considered as a public positive externality, but not a public good. However, the argumentation of information security not being a public good because it is not a single, indivisible good (instead it is the sum of a number of individual firm's or people's decision) and because the solutions to public goods problem might differ from those of externalities (government provision e.g. of national security contra simple interventions to enhance the private market), is valid only if information security is seen as a pure public good – which it is not.

organisation (the appropriability of the value of increased network security is limited and 'free-riding' on the costs of others is possible)[264].

In relation to secure software or information systems development part of information security this means that developing information security properties (e.g. security requirements of confidentiality, integrity and availability) into information or software systems or correcting defects (e.g. by patching) in components used in their development can be seen as causing positive public externalities. When the security of one networked information system is enhanced, also the users and operators of other information systems in that network reap the benefits because there are fewer possibilities for security failures in that network.

From the viewpoint of the overall security of that network, the incentive to invest into the security of one's information systems is too low because the benefits given to others are not considered in deciding about the investment. So, the benefits of improved security are not fully reflected in market prices. When operators, suppliers, or service providers improve the security of their products a good deal of the benefits of this investment accrue not only to their customers but to all those directly or indirectly affected by electronic communication - basically the whole economy[265].

These positive public externalities are relevant also for the component developer, when the products are used in the networked information system development. Adding security features to components used in networked systems enhances the security of the overall network and gives the whole industry a face-lift (customers may consider the products of others also more secure which increases trust as has happened e.g. with the heavy investments made by

---

[264] Similar to public goods, the market fails at the presence of positive public externalities in that it produces too little of the commodities. There is a strong inducement for consumers of the privately provided public good or at the presence of positive public externalities to try to be free riders: they hope to benefit at no cost to themselves from the payment of others.

[265] This is emphasised by the European Commission in its Communication on Network and Information Security: Proposal for a European Policy Approach (COM(2001)298 final, p. 14).

Amazon.com in customer privacy spilling over to other online bookstores), thus creating benefits that the component developer does not consider while deciding about the pricing of the component. Thus a more secure component tends to be overpriced[266].

*Externalities and the quality aspect of security.* 'Public goods' are, however, relevant only in relation to developing security properties Defects (vulnerabilities) in common software products that are used as components in the development of networked information systems are a totally different issue. They cause the whole networked information system to be vulnerable[267]. And since components are increasingly used also in infrastructure critical markets, these bugs make the whole information infrastructure vulnerable. The costs caused by the abuse of these vulnerabilities not only by a hacker or a virus, but also a malfunction or another unintentional act, are not

---

[266] Overpricing is not the only cause for the lack of demand for more secure components. As noted earlier, customers are accepting to errors and malfunctions due to long term industry practice and desire for features and performance instead of security. In addition, the positive externalities caused by the information security investments and raised customer confidence diminishes their willingness to invest into the security of their systems which means that there is less demand for more secure components or add-on security devices.

[267] Even though the diminishment of security related vulnerabilities in a software product used in a networked information system also makes the overall network more secure by diminishing vulnerabilities that could be misused, it is not a positive externality case. The product developed is not vulnerabilities. Instead it is a software component that should have certain quality and security by default. Vulnerability avoidance is not the product even though with patching considered as a sufficient remedy for even a security related defects it seems to have evolved into something like that. But the product is the software that is expected to have quality and security in addition to functionality. So, the spillovers from vulnerabilities being misused (security breaches) resulting in negative externalities are reduced but no positive externalities are produced while diminishing the vulnerabilities. Protecting computer networks from viruses and from hackers reduces the chances that a loss will occur to the agent who takes protection and at the same time reduce negative externalities. Similarly in Kunreuther and Heal, Interdependent Security, p. 21.

paid by the component vendor. When these vulnerabilities are abused and the whole network threatened (not just the systems of component users) they can be seen as externality costs (causing negative public externalities)[268].

> The classification of externalities into positive and negative needs to be restated in here, even though it is a basic tenet in welfare economics. In the case of *positive externalities*, the third-party spillover effects are beneficial — benefits external to the private action that are not captured by either of the transacting parties. In this situation the social marginal benefits exceed the private marginal benefits. For example, if a third party acquires a benefit from a producer's activity without having to pay for it, that benefit will not be reflected (internalised) in the income the producer receives and will not, therefore, be taken into account in making decisions as to how much to produce. As a result, the market fails in that goods exhibiting positive externalities will tend to be under-produced (inefficiently small amount from the society's view is produced) because the producer does not appropriate all of its benefits and sets the market price higher than would if it had reflected the true social benefits.[269] With *negative externalities* there are costs caused to third parties that neither of the transacting parties bear (i.e. social marginal costs exceed private marginal costs). The producer does not bear the full cost of

---

[268] The costs (system recovery and upgrade, loosed business time, costs of dealing with the breaches etc.) are mainly suffered by (are transferred to) the component user (i.e. private customers and developers of an information or software system) or the distributor. Due to the difficulties in proving that just the software of the specific vendor allowed the specific threat took advantage of the vulnerability and the typical (not always valid) liability disclaimers in commercial software, the cost of the security breaches caused by bugs left to a software product are not considered as costs of software development. However, these costs are not externality costs as such, since they are not outside the market system of exchange. The costs are just transferred to the customers. This situation can be corrected by the actions of these private parties – at least in theory and assuming no coordination problems in the actions of the private parties. These costs of security breaches can be seen as private costs – costs that are borne solely by the organization suffering the breach. They should be separated from external cost.

[269] This has been explained by Anthony Ogus in *Regulation*, p. 21.

> his actions and part of it is borne by third parties causing the producer to set the market price lower than if it had reflected the true social costs. Because more will be demanded and hence a larger number of the products will be manufactured and sold at that lower price, goods exhibiting negative externalities will therefore tend to be over-produced (too much output, too much harm). The reason the market fails in the presence of negative externalities (external costs) is that the generator of the externality does not have to pay for harming others, and so exercises too little self-restraint. The externality-generator produces too much output and too much harm because there is a difference between private marginal cost and social marginal cost (the sum of private marginal cost and the additional marginal costs involuntarily imposed on third parties by each unit of production).[270]

In the presence of such external costs the market fails because the generator of the externality (vulnerabilities on software products) does not have to pay for harming others, and so exercises too little self-restraint[271]. The pricing of software does not reflect the possibility and the extent of the damages from security failures associated with the product thus making the pricing of vulnerable software too low. As a consequence, vulnerable software is sold more than more secure ones especially at the presence of asymmetric information typical to security. The development of software components does not reach

---

[270] This understanding is drawn from the economically minded theory of regulation such as Ogus, *Regulation*, p. 18-20, and Baldwin and Cave, *Understanding Regulation*, p. 11. Similar notions are made also in law and economics. See, e.g., Cooter and Ulen, *Law and Economics*, p. 41.

[271] The same reasons apply also to public bads (public negative externality, or simply an 'externality'), which refers to commodities that the consumers do not like as explained by Hal Varian in *Intermediate Microeconomics*, p. 41, as to why markets cannot arise to supply public goods efficiently: the free-rider problem prevents private bargaining solutions to the problem of public negative externalities or public bads and some form of legal intervention is called for. So, at the presence of public bads too much of the commodity is produced together with too much harm. In short, public goods relate to positive public externalities and negative public externalities can be seen as public bads as Robert Cooter and Thomas Ulen explaind in *Law and Economics*, p. 110 and 151.

enough high quality and security from the public interest perspective. And when the cost of poor security is not borne by the source, there is no incentive for the problems to be fixed.

When a defect (vulnerability) in a software – even a well known – is abused e.g. by a virus or a cracker (i.e. there is a security breach), there are cost spilled over to third parties, i.e., others than the software vendor or the customer (often an information system developer). An example of this is a bug in a common software product that makes the information systems using it as a component vulnerable for virus infections; just recall the major virus cases. Lack of security in major software vendors products affect everyone by propagating viruses, reducing bandwidth across the Internet due to spurious traffic, and creating insecure machines that are then used to attack other machines across the Internet[272]. The incurred costs might be significant in relation to infrastructure critical industries[273].

The basic conclusion is that, absent government intervention or other solutions to internalize the externalities, negative externalities are over-provided and positive externalities are under-provided (public bads and public goods). Until these externalities can be turned towards the originator of the vulnerability there will be little economic incentive to do more. Of course, there are other players to consider: the ISPs who manage the broadband connections, the security administrators, and the authors of the malicious code that can most directly be traced to the damage. Sorting out who is responsible for what is a big task, but it is something that needs to be done since right now, the status quo is broken. The true costs of providing effective information security needs to be directed towards those who stand to gain.

---

[272] Shah and Kesan, Incorporating Societal Concerns into Communication Technologies, p. 30.

[273] For example, a breakdown in information security within the telecommunications industry would be likely to cause serious ripple effects within the financial industry, the healthcare industry, and throughout the economy (as well as affecting national defence).

In short, security in terms of vulnerabilities in software components is a negative externality that should be internalized as a cost (negative impact on society) of development in order to be handled properly. Developing secure information systems on the other hand, at the business level of protecting firms assets with more secure systems, is a positive externality (making the whole infrastructure more secure by one firm doing it properly), and thus causes underinvestment into security. This is because it is so difficult to exclude others from benefiting from the work of one party and because one firm doesn't have enough incentives to develop its own security any further than what is necessary to the firm in question.

In the interest of social welfare a number of specific remedies for market failure due to information security externalities could be examined. However, prior to making decisions of specific remedies, more research should be done to measure the true societal costs (i.e., private costs plus the costs of externalities) of security breaches. But this is more of a job for an economist than a lawyer and we can analyze the benefits and disadvantages of the remedies even without specific knowledge of the costs on the basis of assumptions made. But it has to be emphasized that more information on the costs are needed when deciding which remedies to use – even though many are already used even without sufficient knowledge of the true social costs.

## 2.8 Inadequacies in the distribution of security-related information

*Customer incentives and possibilities to acquire security-related information.* Competitive markets can only function properly if consumers are sufficiently well informed to evaluate competing products also on the basis of their quality. Consumers and buyers in general have strong economic incentive to acquire information of products qualities due

to high possibilities for gain[274]. However, the costs of acquiring adequate information on which to make purchasing decisions are often substantial[275]; especially so in information security, where information about security properties and the quality aspect of security is hard to obtain. They are not only difficult to quantify and assess but also time-consuming to evaluate.

In information security markets (both for security features and secure/high quality software) individuals or organisations acquiring and using software for private or commercial purposes operate within an environment in which a great deal is unknown. A lot of it has to do with the characteristics of the commodity protected – information. Information commodities (such as software) in general are not only non-excludable and non-rival to a high degree (the characteristics of public good), they are also highly non-transparent. Mere complexity and scale of software products combined with rapid and continuous changes causes serious problems for customer screening activities in evaluating and testing the quality of the product, producer's abilities and production process. Software and information systems, especially when networked, have become increasingly complex and are reaching a wider market that includes many users with little understanding of the technology or its potential dangers[276].

Most buyers are not knowledgeable about the technical aspects of information security and, therefore, cannot conduct the informed assessment that is needed for sound decision making under the basic

---

[274] This argument is typically made as a theoretical starting point in the economically minded analysis of information regulation like in the comprehensive review of mandatory disclosure from 1981 made by Howard Beales, Richard Craswell and Steven C. Salop, The Efficient Regulation of Consumer Information, p. 502.

[275] This general argument is made by Anthony Ogus, *Regulation*, p. 40, a Professor of Law at the University of Manchester, in an influential book developing an extensive theory of regulation based on legal scholarship and economic research from 1994.

[276] This has been pointed out at the policy level by the European Commission in its communication on network and information security (COM(2001) 298 final, p. 14).

assumptions of economic theories. In addition, the benefits deriving from greater information security or the consequences of inadequate security are difficult to articulate in detail, much less to quantify, which lowers the incentive for the customers to search information.

The information deficit concerns not only consumers that purchase software for private purposes, but also industrial buyers. The higher stakes involved for industrial buyers naturally justify measures to avoid the common information-deficiency problems that consumers face, especially when repeated purchases are made[277]. However, due to the outright unavailability of information on many of the quality aspects of security together with the high costs of acquiring existing information due the lack of methods to gather such information and the underdeveloped state of metrics in the quality aspect of security, not even industrial buyers have sufficient information for sound decision-making. All but the largest and most capable buyer organizations lack the resources or expertise to evaluate the security claims made of a software product[278].

Recall the argument above that the development of the mass market for software products has been accompanied by a shift in systems development and expertise from user organizations to vendors[279]. This means that purchasers have less means and skills to evaluate products. Especially their expertise to evaluate the

---

[277] Of this argument in general see, e.g., the point made by Ejan Mackaay, *Economics of Information and Law*, 146, in an effort to introduce the issues of the costliness of information to the economic analysis of the law in 1982.

[278] Note that the experts used by large buyer organisations obviously have more substantive knowledge than laypeople. Often, however, the practical demands of risk management force experts to make educated guesses about critical facts, taking them far beyond the limits of their data. As noted by Baruch Fischhoff and Jon Merz in The Inconvenient Public, p. 169, empirical studies available in 1994 also suggested that when experts must rely on judgement, their thought processes often resemble those of laypeople.

[279] This is due to the increasing use of COTS components instead of in-house software development (or even information systems development) pointed out, e.g., by Fred B. Schneider in *Trust in Cyberspace*, p. 188 and 198.

development processes is decreased[280]. This is not eased by reverse engineering being made more difficult (source code is kept as a secret, and reverse engineering is prohibited in licences and made possibly illegal in case of products containing security mechanisms designed to protect intellectual property or conditional access) because it makes even more difficult for the consumers to evaluate the vendors claims[281]. Because the gathering of quality information about a certain software product by reverse engineering the code is often not an option, there are fewer provable serious source of quality and security information to serve as a basis for rational choice.

Not surprising, then, is the observation that relatively little information on information security is readily available to purchasers[282]. Paradoxically there is a huge amount of information on network and information security available on the Internet and

---

[280] Since production is separated from consumption, software engineering methods, techniques, and tools are less important to the consumer than is the outcome of their use. That is, vendors are being evaluated by their potential customers on the basis of their products, not their processes as pointed out by Steve Sawyer in A market-based perspective on information systems development, p. 101, and by Erran Carmel and Steve Sawyer in Packaged software development teams, p. 7-19, at the turn of the millennium. Whether this means that COTS vendors are less interested in applying software process improvement methods or even rigorous software development methods is an empirical question that cannot be answered here.

[281] Especially the laws concerning conditional access and intellectual property right protection devices (technological protection measures) are severely hampering researches' ability to verify vendors' security claims and thus increasing the lack of trustful information concerning the quality and security of products. Similarly in Arbaugh, Security: Technical, Social, and Legal Challenges, p. 111. We will come back to this in chapter 4.

[282] This has been recognised also at the policy level in Europe by the Commission of the European Union in its communication on network and information security (COM (2001) 298 final, p. 15) and pointed out by the Computer Science and Telecommunications Board (CSTB) of the U.S. National Research Council (Schneider, *Trust in Cyberspace*, p. 184). CSTB is an independent advisory board for the federal government on technical and public policy issues relating to computing and communications. See the web pages of the National Academies at http://www7.nationalacademies.org/cstb/about.html [21.2.2006].

computer magazines, not to mention professional publications that cover this issue quite extensively. However, the information available on the security and quality of COTS, or any other, software is relatively smaller. Comparisons, studies, and even vendor claims about the quality aspects of security and even of security features of software are still almost non-existent[283]. Even though a variety of incident response centres provide information and different types of guidelines, checklists, best practices, and other useful materials are widely available, the problem for users is to find appropriate information that is understandable, up-to-date, and responds to their particular needs. The existing information is so scattered and difficult to verify that it is difficult to utilise it.

*Seller incentives to provide security-related information.* Sellers can typically provide this information more cheaply because economies of scale are involved. In a competitive market sellers also ought to have a substantial incentive to do this in order to be able to distinguish their products from those of their competitors. There is support in the law-and-economics literature for the hypothesis that, as long as explicit deception is forbidden[284], sellers have incentives to reveal negative attributes of their products. Otherwise consumers will rationally assume that an advertisement will omit a critical piece of information (say, the weight of a notebook computer) only if the value of that

---

[283] As noted by a Canadian Professor of Law at the University of Ottawa Jennifer Chandler in *Improving Software Security*, p. 5, some software vendors even try to suppress the publication of independent product reviews through "anti-benchmarking" clauses included within the end-user license agreement (EULA) for a piece of software. The clauses purport to bar licensees from publishing product reviews or disclosing the results of benchmark tests without the licensor's prior consent. There may be legitimate reasons for these clauses such as the desire to prevent the publication of unfair or erroneous comparisons, but they seem overly restrictive from the perspective of enhancing secure software development.

[284] Most jurisdiction have rules controlling the supply of false or misleading information, i.e., the negative duty not to misinform the other party to a contract (not to lie). Basic examples are provisions on fraud and misrepresentation and its many variants for example in marketing regulation.

attribute for that product is low. Thus, producers of products with quality levels above the minimum will have incentives to advertise this fact[285], and in the limit the market will provide complete information.[286]

Note that this hypothesis underlies an assumption that producers are exogenously informed about the quality and security of their products, i.e., that they know whether or not their products are secure. However, even the COTS software vendor often lacks full knowledge of the quality aspects of security of his product in many cases due to the high complexity of the product and the information concerning it[287]. As has been argued above, testing for security is often bypassed due to budget and scheduling concerns. Not even the vendors know whether their products contain security related vulnerabilities when the product enters the market. The relevant question then becomes, whether vendors are incited both to test and to inform voluntarily.

As Steven Matthews and Andrew Postlewaite show in the terms of economic theory, sellers may not be forced by buyer scepticism to test quality voluntarily and hence to disclose[288]. If the seller can make it *verifiably* known before the market opens that he has not tested quality, then the seller can decide not to test without causing consumers to disbelieve him when he claims to be ignorant. According to Matthews and Postlewaite announcements of ignorance can be

---

[285] Those above the average security would disclose under this hypothesis since otherwise customers would expect the worse of their products. This is expected to lead also others to disclose since they also want to avoid the negative inference of non-disclosure.

[286] This argument has been presented in the Encyclopedia of Law and Economics by Paul Rubin in Information Regulation (incl. Regulation of Advertising), p. 280-281; and by Mark Geistfeld in Products Liability, p. 353. The argument has also been made by Beales et al., The Efficient Regulation of Consumer Information, p. 502, already in 1981. Note that an assumption of the customer's ability to verify the claims of quality and security is underlying this argument.

[287] Similarly by Carl Landwehr in *Improving Information Flow in the Information Security Market*, p. 2.

[288] Matthews and Postlewaite, Quality Testing and Disclosure, p. 329.

credible if testing or disclosing is costly or if the test results are not perfectly informative[289].

As has been argued above, testing procedure for security involves high costs and cannot show the absence of errors; it can only show the presence of errors in the program. This means that vendors can make credible claims of ignorance and continue not to test the program and not to inform without fear of negative inference by the customers about the vulnerability of their software product. In the situation where testing for security is extremely expensive, where it is seen as against the dominating time-to-market principle and does not fully reveal the quality and security of the product, the vendors thus do not have to test for security and to disclose information about the vulnerability of their software in order to dispel cynicism about its products security. They can simply claim that they are ignorant of the issue. In such a situation a firm is not forced to test to dispel scepticism simply because consumers cannot be sceptical.

However, the market may fail to produce adequate information for several reasons even when the vendors do know the quality and security of their products, i.e., they are exogenously informed or do test for security and quality[290]. For example, the dissemination of false

---

[289] Matthews and Postlewaite make this point in Quality Testing and Disclosure, p. 334, with references to similar results made by others in both monopoly and competitive contexts.

[290] Beales et al., The Efficient Regulation of Consumer Information. p. 502-509, and Ogus, *Regulation*, p. 40-41, make this theoretical argument and explain the different possibilities. For the sake of clarity it has to be stated that 'perfect' information, as assumed in traditional economic analysis of markets, never exists in the real world and that the absence of 'perfect' information, from a public interest perspective, cannot by itself justify intervention in markets. The relevant question is whether the unregulated market generates adequate ('optimal') information in relation to a particular area, i.e., the point where the marginal costs of supplying and processing additional information equals to the marginal benefits that are engendered. Even though precise estimations of 'optimal' information are unattainable, it is a useful theoretical tool for identifying situations where the information generated by the unregulated market is likely to be substantially sub-optimal and possibly calls for interventionist measures. (Ogus, *Regulation*, p. 38-39; Beales et al., The Efficient Regulation of Consumer Information, p. 502-503 and 512)

or misleading information and withholding negative information of products qualities, that might be seen as profitable in the short run if the claims are believed and not countered by others, can distort the information provision of a vendor providing both the information and the commodity. Vendors' incentive to distort information provided to the market depends on the prospective gains in terms of increased sales or profit being higher than the prospective losses[291]. The incentive exists especially when, for example, consumers of the product are ill-positioned to challenge the falsification and seek remedies for damages suffered or where they face high costs in doing so[292]. Not only can gains be acquired from the distortion of the information, but also the sale of the product and its value in the market might suffer from the publication of the negative information (e.g., vulnerability of its products) together with the reputation of the vendor itself[293]. Silence about the vulnerability of software or even the distortion of the information to their own favour, e.g., by

---

[291] This basic argument has been made already by Ejan Mackaay in *Economics of Information and Law*, p. 149, in an effort to introduce the issues of the costliness of information to the economic analysis of the law in 1982.

[292] According to Anthony Ogus, *Regulation*, p. 40-41, and Beales et al., The Efficient Regulation of Consumer Information, p. 505-506, especially areas in which consumers purchase a type of product very infrequently may give rise to this problem. Of course the likelihood that purchasers will cease to buy the product due to the discovery of the distortion and encourage others to do so, and the amount of business lost in that case, also affects the calculation. But if the distortion is not discovered by the purchasers, then the losses do not realise. Note that each purchaser does not have to be able to discover distortions in order to alter the vendor's calculation of gains and losses in favour of not distorting the information. Even if most purchasers are ill informed, but at least few customers do inform themselves, and the vendors are not able to distinguish informed from uninformed purchasers, then sellers cannot exploit this ignorance as explained by Ejan Mackaay in *Economics of Information and Law*, p. 149-150.

[293] Mark Geistfeld in Products Liability, p. 353, explains the lack of voluntary disclosure of risk-related information by the consumers' tendency to overreact to negative information about products. Consequently, any seller that discloses risk-related information could cause consumers to believe that its product is unsafe, so high-quality sellers are better off by not disclosing.

promising security and quality that is not actually there, is a tempting option for the vendors.

Especially central for the problem of the security of software is the incentive to hide information about products negative qualities like defects and vulnerabilities[294]. There is a preconception, and anecdotal evidence from the practitioners, that vendors have had a negative attitude towards discovering and reporting software vulnerabilities[295]. Vulnerabilities, especially security related, have been stated to be dismissed by many vendors either due to the fear of negative publicity, or the lack of understanding of the nature of the vulnerability and the risk that it poses, or both. As a consequence, the responses to discoveries of vulnerabilities have largely been seen as defensive, slow and inadequate. Especially the reporters of vulnerabilities have stated that they faced problems when reporting to the vendors[296] and felt that sufficient interest has not been paid to the vulnerabilities.

In information security markets there is not only a reluctance to provide information as to a product's direct negative internal properties (potential bugs and vulnerabilities) but also to make data about actual security incidents and consequences together with actual losses publicly available[297]. The reluctance to make such data publicly

---

[294] The general argument about incentives to hide negative characteristics is usually made in regulatory theory, such as in the comprehensive analysis of governmental regulation done by Anthony Ogus in 1994 (*Regulation*, p. 40-41).

[295] This has been recognised even by the big commercial software vendors. For example, see the comment of the CTO of advanced strategies and policy (at the date of making the comment) at Microsoft Craig Mundie, *Security: Source Access and the Software Ecosystem*, heading "Response Mechanisms".

[296] This is verfied in a quantitative survey study conducted by Tiina Havana, as reported in Havana and Röning in Communication in the Software Vulnerability Process. Problems faced by the reporters include finding the right contact person, to get a response to a report etc.

[297] This has been identified, e.g., by Fred B. Schneider, *Trust in Cyberspace*, p. 184, in an influential study from the end of 20th century. See also Landwehr, *Improving Information Flow in the Information Security Market*, p. 2. One of the key findings of the Tenth Annual CSI/FBI Computer Crime and Security Survey 2005 is that the percentage of organisations reporting computer intrusions to

available is intended to minimize the public perception and awareness that systems are vulnerable or have been breached. Potential risks are underplayed, for fear of losing customers. Competition over the degree of vulnerability may decrease total purchases of the product rather than help any particular manufacturer to obtain greater sales.[298]

## 2.9 Asymmetry of security related information

The above mentioned possible market failures, i.e., inadequacies in the distribution of security-related information, may occur in the production and sale of information in relation to quality of software and its security features. However, informational imperfections may also prevent the underlying product markets from working properly. One such situation is when sellers know more about a product than do buyers, or vice versa (asymmetric distribution of information)[299].

---

law enforcement has continued its multiyear decline (Gordon et al., *CSI/FBI Computer Crime and Security Survey*, p. 2). The predominant reason given for not reporting intrusions to law enforcement cited as being very important (by those indicating that their organizations would not report an intrusion to law enforcement) was the perception that the negative publicity would hurt their organization's stock and/or image. As also pointed out by Gordon et al. in *CSI/FBI Computer Crime and Security Survey*, p. 20, this is consistent with the research by Campbell et al. in The Economic Cost of Publicly Announced Information Security Breaches, p. 431-448, from 2003 that found reports of security breaches of personal data having a special adverse effect on a firm's stock price.

[298] Schneider, *Trust in Cyberspace*, p. 184. Cass R. Sunstein, The Functions of Regulatory Statutes, p. 52, makes the same argument in relation to all manufacturers that have poor incentives to provide information about hazardous products.

[299] The case where the informational advantage is on the consumer side can give rise to *moral hazard* in the enforcement of contracts, which means that parties misuse their information advantage. This arises particularly in insurance markets, where insured parties have a tendency to be more careless due to the awareness of compensation in case of accident. This leads bigger accidents and bigger insurance claims. This will be ignored in the analysis.

In information security markets and in markets for secure software producers sometimes know which products are secure, but consumers cannot tell. A relevant information asymmetry arises because information concerning quality or security is more costly to supply and process than information concerning price or quantity. Prices are calculated by reference to objective criteria (currency) and, in general, are easily communicated. Qualities, such as security, are to some degree subjective and, particularly in the case of professional services and technologically more complex commodities like software, may not be discoverable by pre-purchase inspection[300].

In a famous paper, The Market for "Lemons", mentioned in the Nobel Price in Economic Sciences won in 2001, George A. Akerlof demonstrates how asymmetry in the search costs of price and quality information can lead to seriously detrimental consequences[301]. Failure arises in a lemons market because only low quality items are sold, even though consumers would be willing to pay high prices for high quality items. This theory predicts that this asymmetry in information will force more secure products out of the market if secure products will sell for no higher price of than vulnerable ones, secure products will be more expensive to produce, and consumers will not be able to know the difference. This process is called *adverse selection*.

The three conditions necessary to generate a lemons market (adverse selection type of market failure) are present to certain degree in information security and secure software markets[302].

---

[300] This means that potential buyers do not have reliable information about whether the software they intend to buy meets their specific needs and is of appropriate quality and security. They first have to buy the product and can rate its quality and security only after having deployed it. At this time, the software producer already has made his profit.

[301] Akerlof, The Market for "Lemons", p. 488-500.

[302] These conditions are presented in a general form, e.g., by Paul H. Rubin in his article in the Encyclopedia of Law and Economics, Information Regulation (incl. Regulation of Advertising), p. 278.
The underlying condition that alternatives exist in the market is assumed and extensively experienced in the software component market even in relation to operating systems. Variety in products and vendors is increasing.

First, consumers are unable to determine quality and security before purchase because software as an information product is essentially an experience good, as pointed out by Carl Shapiro and Hal Varian[303], similar to the complex information systems that are possibly networked[304]. Its quality is an experience characteristic[305], and the quality aspects of security even a credence characteristic[306]. Second, more secure and higher quality products cost more to produce than lower quality, as has been noted several times above. The third, and last, condition according to which there cannot be a credible way for a firm to guarantee quality requires a wider analysis.

The importance of this analysis stems from the notion that to the degree the last of the three conditions is met, i.e., whether or not there are credible ways for firms to guarantee quality aspect of security in

---

[303] Shapiro and Varian, *Information Rules*, p. 5.

[304] Economic theory distinguishes three types of goods: search, experience and credence goods. While the quality of some products ('search goods') can be determined prior to purchase and asymmetric information on a good's characteristics can be eliminated before consumption takes place by paying a search cost, others ('experience goods'), including almost all types of services and technologically more complex products, can be evaluated only in the process of receipt, use, or consumption. In some cases ('credence goods'), the quality is known only years later or cannot even be established after consumption at all. See, e.g. Ogus, Regulation, p. 132-133; den Hertog, General Theories of Regulation, p. 228-9); Noll, Comparing Quality Signals as Tools of Consumer Protection, p. 228); Rubin, Information Regulation (incl. Regulation of Advertising), p. 277.

[305] It is often difficult to observe the quality of a software product before purchase. As Juergen Noll notes in Comparing Quality Signals as Tools of Consumer Protection, p. 228, that identifies product quality as experience property.

[306] Whereas some dimensions of quality, such as reliability and durability, can be objectively determined, others involve a high degree of subjectivity and some are visible only after some time. Some of the quality aspects of security cannot be determined even after consumption or can be assessed only with highly sophisticated technical help, since security related vulnerabilities typically are not visible in the functionality of the program and testing cannot show the absence of vulnerabilities (it can only show that there are defects). Thus, parts of the quality and security characteristics of products belong to the credence category.

their software products, then the market mechanism may break down. This will happen because no firm will be able to convincingly promise high quality items. As a result consumers cannot be sure of obtaining the higher quality and so will not pay the higher price for quality items. Thus, even though consumers would be willing to pay a higher price in order to obtain quality, there will not be an effective way in which this desire can be satisfied.[307]

*Convincing communication of quality and lemons market.* The lemons problem identified by Akerlof exists only if firms cannot convincingly communicate to consumers the level of quality in their products[308]. If firms can produce high quality products and convince consumers that they are doing so, then the market failure disappears. In most cases informational asymmetries can be corrected by the mechanism of voluntary exchange and, as Hugh Collins notes, "[t]he absence of any regulation of quality would therefore not lead to the widespread supply of defective products and shoddy services"[309]. However, whether this happens in relation to the security of software requires a more detailed analysis.

There are several means by which vendors can convincingly communicate the level of quality in their products and thus correct the market failure by themselves. Typical examples are reputation, advertising and other voluntary information provision by the vendor

---

[307] The underlying condition that alternatives exist in the market is assumed and extensively experienced in the software component market even in relation to operating systems. Variety in products and vendors is increasing.

[308] Note that this is the sharp distinction between the above notion of sellers having strong incentives to provide adequate information. This incentive exists only if there is some way for consumers to check on the claims of the sellers. If the lemons problem can be solved, sellers of higher than average quality products will have incentives to reveal information about their products quality. Consumers may then assume that any product which does not disclose quality is of below average quality, and the informational problem is solved. (Rubin, Information Regulation (incl. Regulation of Advertising), p. 281).

[309] Collins, *Regulating Contracts*, p. 288.

(e.g. in terms of test results), and warranties to guarantee the quality of the product[310].

It has to be stressed that the means of vendors to communicate the level of quality in their products (such as reputation, advertising, warranties etc.), and thus to correct the informational market failure voluntarily do not apply to credence characteristics of products and are unnecessary for search goods[311]. This means that these measures are necessary and effective only in relation to the experience characteristics of software security. This already limits the applicability of these remedies in relation to most security related vulnerabilities that are at the core of this study, due to them being credence characteristics of software.

Advertisements constitute the most obvious method of communicating quality information but, as already discussed, under the situation where customers lack the means to verify the correctness of the claims as is the case in relation to software vulnerabilities to a large degree especially due to credence characteristics of the vulnerability information, the opportunity to disseminate false or misleading information or withholding negative information seems profitable in the short run[312]. Only when the quality and security information is easily controllable, which is not the case with the quality aspect of security as discussed above, can the customers expect to get truthful quality and security information. Thus, advertising and other general information sharing by industry is not a likely means

---

[310] Guarantees, reputation and licensing were identified already in 1970 by Akerlof in The Market for "Lemons", p. 499-450, as institutions counteracting the effect of quality uncertainty. Anthony Ogus in *Regulation*, p. 133, explain these mechanisms in general terms in his extensive analysis of governmental regulation from the economic perspective.

[311] Noll, Comparing Quality Signals as Tools of Consumer Protection, p. 228.

[312] The general argument about the false claims as a form of information failure is made, e.g., by Beales et al. in The Efficient Regulation of Consumer Information, p. 505-6.

of correcting the informational asymmetries related to the quality aspect of security[313].

General voluntary disclosure of security information is further hampered by the possible spillovers which result in positive externalities for the industry as a whole. This is so especially in relation to the customer confidence aspect; enhanced customer trust in transacting with a particular firm also expands the overall market size within the industry[314]. Seller-provided information creates externalities that can lead to an undersupply of general information. In particular, advertising that provides positive general information about all brands in a product class benefits every brand, not simply the one generating the information. In such as case, the disclosing firm's competitors will share in the benefits as free riders[315]. Advertising that provides negative general information about a product class is likely to reduce the sales of each firm[316] and possibly benefit the sales of substitute products, thus reducing the incentive of any single seller to provide this information.[317]

The willingness of a firm to spend money on advertising as such, without necessarily offering information, can in itself also be a signal of the quality of the product. The economic argument is that advertising is worthwhile only if it leads to repeat sales for experience

---

[313] Not even competitors might have sufficient incentive to intervene in unfair marketing practices. This is especially so when they share the same negative attribute (vulnerability of software) or are subject to externalities like the benefits from corrected customer beliefs that have to be shared with other competitors and are thus inadequately internalised by the intervening competitor, or like the increased customer belief that a proportion of security claims are false that harms the industry in general. (Beales et al., The Efficient Regulation of Consumer Information, p. 506)

[314] Esther Gal-Or and Anindya Ghose make this argument in *The Economic Consequences of Sharing Security Information*, p. 3.

[315] Beales et al., The Efficient Regulation of Consumer Information, p. 527.

[316] I.e, competition over vulnerability reduces the sale of the whole COTS industry, instead of enhancing the sale of certain software brand inside it.

[317] This argument in general is made, e.g., by Beales et al., The Efficient Regulation of Consumer Information, p. 503-4.

goods. Firms can expect repeat sales only if the product is of sufficiently high quality. Therefore the investments in advertising signal to the market that the firm expects repeat sales because it believes that its products are of high quality. Same line of argumentation applies also to investments in establishing trademarks and brand names, and also in physical assets, such as signs and décor.[318]

Reputation, which is largely connected to the above mentioned investments on capital and advertising, that will be lost if the firm goes out of business, as such can serve as an indicator of quality. A firm selling low-quality products at high-quality prices will soon acquire a bad reputation and be excluded from the market. Consumers may, over time, also accumulate trust in the quality of a particular firm's output or a particular brand name. With the desire to preserve goodwill, it will be in the interest of the supplier or brand manufacturer to maintain quality[319]. But the reputation mechanism is workable only if the product is an experience good, i.e. that buyers can find out the quality and security of the product after buying it.

As has been argued, this is not the case with the credence characteristics like quality aspect of security. Even when the buyers can find out the quality and security of the product after buying it, reputation adjustment only rewards quality upgrading with a time lag, which means that firms will not provide so high quality as with under perfect information. Therefore, despite being necessary in certain surroundings, reputation is an unreliable indicator of the quality aspect of security.[320]

---

[318] Advertising as a signal of product quality has been analysed in the law and economics literature, e.g., by Paul H. Rubin in Information Regulation (incl. Regulation of Advertising), p. 278-279, and Juergen Noll in Comparing Quality Signals as Tools of Consumer Protection, p. 229.

[319] This does not apply in relation to one-off transactions; but even here, reputation may have a value when recommendations are made by friends or relatives as noted by Anthony Ogus in *Regulation*, p. 133.

[320] See Noll, Comparing Quality Signals as Tools of Consumer Protection, p. 229-230, for the general argument.

Contractual terms, such as product warranties or even money back guarantees, maybe more reliable signal of quality. They partially indemnify the buyer against the possibility that lack of information leads to making a wrong choice. In addition to this insurance effect, such contractual terms may as such signal of the quality of the product since warranties and money back guarantees are cheaper to provide if product failures seldom occur.[321]

However, warranties are almost inexistent in software markets. Software vendors assume no liability and try to avoid giving any warranties (not that even of merchantability, fitness for purpose or any alike) by standard licensing provisions[322]. The justifications wary but the essential point is that unlike traditional commodities, defects in software are likely to exist in every copy sold, thus making the compensation or repair especially expensive. Disclaimance is a routine despite the threat of suspicion towards vendors lack of care of software quality and security being fed and the confidence into the quality of the software products possibly being eroded by disclaimers of warranties with which the products may well comply. Even in cases where warranties still exist despite of the software licences, the incentives for their enforcement are still reduced since the likelihood

---

[321] This argument is made, e.g., by Ogus in *Regulation*, p. 133 and Beales et al. in The Efficient Regulation of Consumer Information, p. 511.

[322] Note that industrial buyers have at least a theoretical possibility to buy better guarantees and to negotiate requirements for secure development into the agreement with the consequence of needing to pay more for the software at the same time. Without any empirical facts one can only assume that the price charged for the software sold with security guarantees would be extensively higher than its list price. Whether industrial buyers would be likely to buy better guarantees if they do not perceive the vulnerability issue as an important factor in purchasing decisions is a matter of empirical study that still, to my knowledge, remains unanswered. In addition, high transaction costs in mass-marketed COTS software products at least diminishes, if not prevents altogether, also the possibility for industrial buyers to settle warranties individually.

of successful litigation is decreased and the claims often cannot be enforced because of the global nature of the market[323].

In addition to trials, typically used in marketing of more traditional goods, there are special ways for the vendors to provide information about the functionalities of the program. The logic in shareware, allowing interested persons to download the full program, is that if they like it they would send its author some money, and perhaps in return you get a manual, access to support, and/or an upgraded version[324]. The public beta is a time-limited (or bug-ridden, or otherwise restricted) version of the product. It allows users to investigate the properties of the public beta version to figure out whether the product is worthwhile. But to get the permanent (or the less bug-ridden) version, they have to pay. A similar method is the free provision of lesser version of programs and making advanced version chargeable.[325]

These methods enable the possible purchaser to familiarise with the experience characteristics before purchase and lower the costs of finding information about the software. Even though competition through distribution of lesser versions of the ultimate product as such is a relatively benign development, they are not providing information about the security (excluding observable security features). In addition, as argued above, the first-to-market competition (which partly explains

---

[323] It is worth pointing out that in cases where warranties are based on extensively on law, as in consumer law, their informational value for the customer and the signalling value for the producers are diminished (Noll 2003, p. 219-231).

[324] Even though this try-before-you-by characteristic makes software more transparent, it also easily leads to users not paying for the software.

[325] These information provision mechanisms have been analysed by J. Bradford DeLong and A. Michael Froomkin in Speculative Microeconomics for Tomorrow's Economy, heading "The Market for Software: Shareware, Public Betas and More". Newest development in the correction of the transparency problem is open source that possibly can also be beneficial for security. Note that its main function is not to provide information for users since only a possibility to derive security information from the source code is enabled. Instead it constitutes a method of developing software and ideological approaches to software markets.

this development since these methods to provide information can be ways to introduce products to the markets earlier) can mean that the security of these lesser versions is not necessarily increasing. At least, these methods to provide information do not correct the informational asymmetries in relation to quality and security of the 'final' product. They are efficient only in separating the features that the product contains – not their quality or security. In addition they can possibly be detrimental for security since they not only provide no incentive to develop secure and high quality software but can also provide opposite incentives.

In sum, it would be irrational behaviour of customers to prefer to buy more expensive software because vendor's claims of higher quality and security cannot be verified. The claims of quality and security may be right when made, but the customer cannot judge them. According to economic models of customer behaviour, they should and will decide to buy the cheaper product that may or may not be of lower quality and security. So, even though consumers would be willing to pay higher prices for high quality and more secure systems and components[326], in lemons market only low quality items are sold.

Making purchasing decision for secure systems under imperfect information increases the level of uncertainty regarding the benefits of using more secure systems. This is a disincentive to invest in security and distorts the market. Because it takes significant time and energy to extract the information needed to permit rational decisions, the buyer commonly faces a choice between new, unevaluated products or systems that perform better, but have uncertain security properties, and older products with better known properties but

---

[326] This is a matter of controversy. People tend to assert that they are concerned with privacy and security and are willing to even pay for it. However, their behaviour does not always agree. For wider discussion, see the analysis of Alessandro Acquisti and Jens Grossklags in *Losses, Gains, and Hyperbolic Discounting*, from 2003. However, the general assumption goes in favour of buyers' readiness to pay for increased security and higher quality. Many of the arguments by security professionals that consumers lack the interest in investing into information security are explained by the asymmetry in the information or the simple lack of it.

poorer performance. Due to the uncertainty, buyers prefer to purchase greater functionality rather than to invest in improved security.[327]

In addition to the above provided historical explanations on why buyers are so approving in relation to the vulnerabilities in software[328], a partial explanation is provided by an informational market failure that affects the relevant commodity market called misperception[329]. Security related vulnerabilities, typically a credence characteristic in software (i.e. they cannot be appropriately assessed even after consumption or only a long time after that when, e.g., an attack occurs), and especially their exploitation or seriousness tend to be misestimated due to the small probabilities involved. This results in buyers believing they have made optimal choices, while objectively by their own standards they have not.

Due to this misperception the consumer demand is insufficient to persuade the vendors to assume such an uncertain cost that might be high due to vulnerabilities existing in every copy, even though the vulnerabilities are a statistical fact to the vendors as they deal in large numbers. As a consequence, the demand in the market, which reflects the aggregate of individual preferences, will not represent the optimal quality of the software.

The reason why buyers as a group have not for long insisted on having information on the problem or on shifting the risk to the producer is that the experience feedback on which most purchasers rely to decide whether risks are worth bothering with suggest that they are not. The awareness of vulnerabilities and the subsequent costs have for long not been high enough to warrant action. In addition, the reason why the collected information on the disruptive consequences and high costs of vulnerabilities, which there is

---

[327] Schneider, Trust in Cyberspace, p. 184 and 251. Limited actual experience with losses from security incidents also tends to discourage investments in trustworthiness, even though there currently are more estimates on the actual and potential losses available.

[328] See footnote 238 and accompanying text.

[329] On misperception in general, see MacKaay, *Economics of Information and Law*, p. 166-169.

significant reluctance to report, has not lead to autonomous market correction either through safer products or through independent accident insurance is that the cases have been considered as risk of doing e-business. This is especially so because protective action by the purchaser, e.g. in terms of firewalls, virus-scanners or other defensive or offensive information security products, significantly can reduce the risk.

Partly consequently to these, software vendors have a higher incentive to offer new features than greater security[330]. The benefits of the new attractive features are visible to the purchasers, but the risks are not due to the lack of information and the capability to understand it. There is no need to compete on security grounds, since the consumers do not know to demand it or do not know to evaluate secure and vulnerable products. Since the buying public has no way to differentiate real security from bad security, the way to win in this marketplace, just lightly overstating, is to design software that is as insecure as you can possibly get away with. Secure systems, that are harder, slower, and more expensive, both to design and to implement, do not have a chance. Not only are new features overriding increased security and causing security stay at the same level, but also the quality and security of products is not enhancing (especially due information asymmetry) as the Akerlof's theory of 'lemons market' suggests.

It has to be stressed again in here that merely finding that buyers lack certain information does not imply that governmental or any other type of regulation is directly needed. Even though vulnerability information is not readily available from another source and firms do not have their own incentives to disclose whenever disclosure would be useful, which means that the case for required disclosure

[330] As noted also at the policy level in Europe by the Commission in its communication on the proposal for a European policy approach to network and information security (COM(2001) 298 final, p. 14).

is strong[331], the information might, however, be so expensive to produce that it is not efficient to provide it or it might be of little value to consumers[332].

Information about the security and quality of software is, however, of high value not only for organisational buyers but also for individuals. Lack of it can lead to serious hazards for safety and produce significant economic losses as the virus incidents alone have shown. The fact that software vulnerabilities make the whole information infrastructure vulnerable and involve negative externalities, thus eroding the trust required for the e-commerce and e-administration to bloom, is enough to show the societal desirability of more secure software. The social value of vulnerability information exceeds the costs of information[333].

Since it is socially desirable to have the information about the quality aspects of security available in the markets despite the fact that it might be expensive to produce this information, the relevant question is which of the parties have better access to vulnerability information by the means of testing and whether or not that party is incited to acquire vulnerability information and not only to disclose it. In the case of software security, there is no doubt that the vendors have a better access to quality and security information by the means of testing. Customers' lack of experience and inability to perform many of the testing procedures has already been shown above. And

---

[331] The need for requiring disclosure in general, as explained by Howard Beales, Richard Craswell and Steven C. Salop in The Efficient Regulation of Consumer Information, p. 527, in their comprehensive review of mandatory disclosure from 1981, depends on the completeness of the total information environment and seller's incentives to disclose voluntarily.

[332] This point is made more generally by Beales et al. in The Efficient Regulation of Consumer Information, p. 510

[333] In situations where only the effectiveness or operating costs of the software product depends upon quality, it is not clear that welfare is greatest in a regime where consumers are informed before they purchase as argued more generally in economic theory by Steven Matthews and Andrew Postlewaite in Quality Testing and Disclosure, p. 330. Releasing quality information may only cause the price of a product to be positively correlated with its quality.

it is not just about access to source code. It is about access to the scarce and expensive experts who can test for security and about the shift of expertise in software development away from user organisation. This makes the case for disclosure rules that also encourage testing by the vendors favourable.

In short, severe informational asymmetries disrupt markets so much that a social optimum in secure software development cannot be achieved by voluntary exchange. In such a situation, an intervention outside the market (e.g., by a government or by any other regulator) might be an option as regulatory theory informed with the economic analysis of the law suggests[334].

---

[334] See, e.g., Ogus, *Regulation*, p. 133; Beales et al., The Efficient Regulation of Consumer Information, p. 511-512; Cooter and Ulen, *Law and Economics*, p. 43. Regulatory theory explains that the case for disclosure regulation is weak where, as a consequence of market competition, producers have an incentive to disclose relevant information voluntarily. This is the case especially in relation to 'search goods' or 'search characteristics' of goods, particularly if there is a significant number of discriminating consumers purchasing at the margin. In relation to 'experience goods' the incentive is weaker, but even here, producers will often have large investments in a brand name and they will not want to lose the goodwill attached to it. The protection of trademarks, for example, is based largely on the belief that distinctive trademarks make it easier for consumers to develop a body of information (largely from their own experience) about individual brands, thus rewarding sellers whose brands achieve a good reputation and penalizing those sellers whose brands do not as argued by Beales et al. in The Efficient Regulation of Consumer Information, p. 493. Voluntary disclosure is least likely in relation to 'credence goods', because for reasons of competition, sellers will not be motivated to refer to negative qualities, particularly when the risk of injuries is so remote that consumers will be unable to establish a tort claim for product liability. (Ogus, Regulation, p. 134)
In relation to search characteristics, consumers can immediately determine if the good possesses the advertised characteristic, and cannot be deceived. Moreover, since this is so and firms understand that it is so, there is no incentive for deceptive advertising with respect to these characteristics. For inexpensive goods, there is little cost to deception with respect to experience characteristics. The consumer will be deceived at most one time with respect to such goods, and therefore in general losses will be small. This means that regulators ought to concentrate on relatively expensive experience goods and particularly on credence goods. (Rubin, Information Regulation (incl. Regulation of Advertising), p. 278)

# 3 The way regulation affects behaviour

In the earlier chapter, the point of view has been that of computer and information system scientists, software engineers and practitioners of information security and especially the secure software development part of it. The main concern has been on what information security is, how it operates and is affected by the business models and market structures. Understanding of the business models, the ways software and information systems are developed together with the environment where it is done is crucial if we want to understand how it might be possible to influence the behaviour of the actors.

Now we can turn to the issues how regulation guides behaviour. The effort is to try to slightly open the veil covering the influence mechanisms of regulation. The intention is not to give a comprehensive overview of the ways regulation affects behaviour[335]. The uncertainty in the guidance mechanisms largely remains[336]. In order to understand the regulatory capability of an instrument one needs to understand (at some level) the mechanisms by which the instrument causes effects. Only the main ways regulation affects behaviour are made explicit in here; not all of them. There might be,

---

[335] A deeper analytic isolation of the ways regulation affects behaviour would require one to make a detailed and systematic study into the wide discussions in a variety of disciplines that have an interest in the ways regulation affects behaviour. Much has been said about the influence mechanisms of regulation or rules in general from the perspectives of economics, sociology, psychology, philosophy and legal theory. Going through all this literature is, however, outside the scope of this study and hopefully subject the forthcoming studies. The approach is somewhat eclectic by necessity.

[336] The uncertainty exists even for the guidance mechanisms of the most widely studied tool of regulation, i.e., the law, as has been emphasised, for example, in the Nordic discussion on legislative theory by Hellner, *Lagstiftning inom förmögenhetsrätten*, p. 164.

and certainly are, a variety of other influence mechanisms and the ones presented here could be conceptualised differently.

This normative analysis of the ways regulation affects behaviour and of the instruments used is instrumental for the main purpose of this study, i.e., the analysis of the capacities of two most widely called for regulatory instruments for solving problems in secure software development. At the same time it is the second main line of argumentation in this study. The heuristic devices for the analysis of the ways regulation influences behaviour are provided in this section.

## 3.1 By providing reasons for action

According to the law and economic scholars' correct criticism, traditional legal analysis has little means beside intuition and available facts to study the effects of laws[337]. This is partly because the doctrinal legal education grounded in descriptive legal scholarship gives a limited understanding of the ways the law influences behaviour[338]. Not even the normative guidance provided by law is clear and interest in the

---

[337] Cooter and Ulen, *Law and Economics*, p. 3.

[338] At the same time this explains why lawyers tend to be overly optimistic in their belief about the influence of the law on behaviour (of this optimism, see Mathiesen, *Rätten i samhället*, p. 158). Claes Sandgren, Professor of Private law at the Department of Law of the University of Stockholm, provides an explanation in Om empiri och rättsvetenskap, p. 739-740 (with reference to Torstein Eckhoff (1985) Empiriske metoder i rettsvitenskapen, in *Rationalitet och empiri*, Juridiska fakulteten if Stockholm, Skriftserien nro. 6, p. 32-35): since lawyers generally tend to overemphasise the norm-orientation of people and, as a consequence, to build their analyses on unreliable premises and preconceived impressions on the effects of laws on behaviour, they are keen to believing that laws have their intended effects. In other words, traditional jurisprudence concentrates on deliberate law-making and does not put emphasis on the effects of laws that are not aimed at. In doing so it overestimates the power of law as a means for social engineering as Mark Van Hoecke points out in *Law as Communication*, p. 4, with reference to Summer RS (1977) Naïve Instrumentalism and the Law, in Hacker PMS and Raz J (eds.) *Law, Morality and Society. Essays in Honour of HLA Hart*, Clarendon Press, Oxford, p. 119-131.

issue is scarce outside legal theory[339]. This emphasises even further in relation to the wider aspect of regulation.

The inadequacy of the hierarchical, top-down commanding model, according to which the objects of regulation follow the orders of the regulator (possibly backed by sanctions) in a uniform and predictable way is obvious despite the fact that it is still widely used in the studies on the effects of legislation and a general assumption made by legal scholars when they consider the effects of laws[340]. Behaviour modification that is associated with the activity of ensuring compliance with rules through strategies of enforcement (i.e., being accompanied

---

[339] The way the law influences behaviour of ordinary citizens has not been subject to great interest even in legal theory. The concentration is typically on judicial decision-making, especially by judges, and on the way regulation influences their decision-making. Even though many of the consideration are derived from general theory of action that is applicable more generally, legal theory has mostly concentrated on applying it to judicial decision-making. Not even the creation of law by the legislator has been of much interest in legal theory due to the separation of law and politics in the mainstream legal thinking as Luc Wintgens, Legisprudence as a New Theory of Legislation, p. 3, argues in a recent effort to apply the tools of legal theory to legislative problems. Extreme concentration on adjudication in legal scholarship has been criticised already in the 1970s as Luc Wintgens points out in Legislation as an Object of Study of Legal Theory: Legisprudence, p. 9.
The understanding of the ways regulation affect behaviour is crucial, however, also for legal scholarship. According to the basic premise of the economic approach to law, which is also shared by the scholars in other behavioural sciences like sociology and political science, the nature of legal institutions cannot be understood by limiting the consideration to the legal arguments alone. It is essential to consider as well what effects those institutions have on society and what reactions they will evoke from the citizens as result. This argument has also been made in relation to the economics of information by Ejan Mackaay in *Economics of Information and Law*, p. 4.

[340] Even though this assumption of mechanical obedience of the law lies in the background of many positivist approaches to the study of law, it has to be noted that not all positivist general theories of the law presume this. As pointed out by Ruth Gavison in Comment: Legal Theory and the Role of Rules, footnote 103 and accompanying text, positivist legal theoreticians have held a variety of positions on the obligation to obey the law. For example, Joseph Raz in *The Authority of Law,* p. 233-250, argues that there is no general obligation to obey the law.

with deterrent effects) goes a little further, but not much. A better understanding is needed in order to study the effects of regulation.

But neither is economics *the* silver bullet that explains the ways regulation affects behaviour. When analysing the effects of regulation the standard economic principles and assumptions about human behaviour used in the economic approach to law have to be lowered to the same level with arguments from other disciplines. The argument is not that the law and economic type of behaviour prediction is not valid[341]. Empowered with the simple assumption that legal consequences (e.g., sanctions, rewards, liability, remedies) affect the implicit prices of actions[342], and accompanied with the mathematically precise theories (price theory and game theory) and empirically sound methods of economics (statistics and econometrics), the law-and-economics studies have been able to predict how people respond to changes in laws. It might be true that the economics provides a single best, even moderately comprehensive, theory of regulatory phenomena (especially legal) and human behaviour from which verifiable propositions can derived and that empirical studies tend to support[343]. As a method it is the most advanced and tested way of predicting the effects of regulation on behaviour and it has provided several important insights about the ways regulation influences behaviour.

---

[341] To treat the law as an incentive for changing behaviour (implicit prices) is the core of law and economics and the most praised, and the least contradictory, part of the economic analysis of the law. Similar analysis can be applied, and have been applied, also to the regulation more widely. The arguments developed in these studies, which are collected in scholarly books on regulation, are used to analyse the effects of specific instruments when available.

[342] For example, the effects of sanctions and private law remedies on behaviour are analysed like the effects of prices.

[343] This is emphasised by Ejan Mackaay in his analysis of the economics of information and law (*Economics of Information and Law*, p. 4 especially endnote 10 with references to Posner). The problems of sociological studies in providing even a somewhat comprehensive model of the effects of laws has been addressed by one of the most influential Nordic sociolegal scholars Vilhelm Aubert in *Rettens sosiale funksjon*, p. 164.

But we are not concerned with the prediction of behaviour as such in this study. The analysis concerns more widely the ways regulation influences behaviour and the possibilities it has to guide behaviour related to the specific topic of secure software development. The quest for the ways regulation affects behaviour is even more interdisciplinary than law and economics. Economics is just a heuristic device in analysing the ways regulation affects behaviour, not a comprehensive method, as Hugh Collins has emphasised in his regulatory analysis of contracts and contracting[344]. In order to be able to comprehend the efficacy of regulation on this specific topic, the disciplinary basis has to be widened[345]. Overall, results and observations utilised in this study come from disciplines like political science, sociology, economics, policy analysis (especially implementation and evaluation

---

[344] Collins, *Regulating Contracts*, p. 7. This means that I do not tie myself down to normative law-and-economics which holds that legal norms must be formed in a way that maximum economic efficiency is fulfilled. The analysis of the effects of regulation is not the sole factor from which behavioural norms or justifications for their improvement are derived. However, in the descriptive sense law-and-economics can produce information about the effects of regulation that is useful in the evaluation of positive valid norms and different regulatory choices. A leading Finnish private law scholar Mika Hemmo in *Vahingonkorvauksen sovittelu ja moderni korvausoikeus*, p. 25-32, makes similar considerations in analysing the reliability of the information provided by law-and-economics research in revising tort doctrines.

[345] Economics as a discipline has admitted its need to widen the basis for analysing human behaviour. The typical account of the standard economic principles according to which the economic approach to law analysis legal rules, i.e., that all human behaviour can be viewed as involving participants who maximise their utility from a stable set of preferences and accumulate an optimal amount of information and other inputs in a variety of markets, is currently being modified under the banner of behavioural law and economics in order to be able to "model and predict behaviour relevant to law with the above mentioned tools of economic analysis, but with more accurate assumptions about human behaviour, and more accurate predictions and prescriptions about law" (Jolls et al., A Behavioural Approach to Law and Economics, p. 1474 and 1476). With the insights from other disciplines about actual human behaviour currently being internalised into the economic approach, it is likely to be even a more useful tool for regulatory analysis.

studies), legal theory, philosophy, management studies, software engineering, and information systems science[346].

However, the path taken in economic analysis is correct; the concentration on the decision-making of small groups, such as individuals and firms. But there is no reason to adhere solely to the price theory or even to economics more generally. It is a widely accepted assumption in societal studies in law and regulation, that the effects are always caused by the actions of the objects of regulation (individuals, corporations, associations, institutions) that make the final decision to react (or not) to a specific regulation, or to redirect their actions or beliefs in the intended way (intended by the regulator)[347]. Regulation affects through the decisions and choices, either conscious or unconscious, made by its objects.

This means that a regulator can try to alter the behaviour of the addressees only by influencing their choices or desires in a decision-making process, in their practical deliberation (reasoning) on what to do or not to do. Regulation can only provide reasons for action

---

[346] The necessary multidisciplinary approach in regulatory studies in general is emphasised by Baldwin and Cave in their basic study book *Understanding Regulation*, p. 1. Despite the usage of interdisciplinary material in this study, i.e., the adoption of a more external point of view to regulation that is more interested in the interaction between the regulation and the social practice that it seeks to alter, the internal aspect of the law that the legal scholarship employs plays an important role. The internal aspect to law and regulation is necessary to see how those functioning inside the regulatory system think about the social practices under regulation and how regulatory objectives are defined and implemented. This dimension is important for any critical assessment of the potential efficacy of regulation. The internal functioning of the regulatory system and the considerations of those functioning inside centrally shapes the effects regulation can have and influences its efficacy.

[347] This has been presented as a starting point in studies about legislation or theory of legislation such as Tala, *Lakien vaikutukset*, p. 16 and 286, in basic textbooks on sociolegal studies such as Mathiesen, *Rätten i samhället*, p. 30, and in political studies and especially studies on policy tools such as Schneider and Ingram, Behavioural Assumptions of Policy Tools, p. 513-514, or Daintith, Law as Policy Instrument, p. 28. This is also the typical approach of legal philosophy and legal theory as explicated, e.g., by MacCormick,On Legal Decisions and their Consequences, p. 254, when he reminds us that "the law and rulings in law … are grounds for choice by people, and how people will choose to respond is always in some degree an open question".

or inaction (i.e., incentives); an individual can always choose not to comply with the provision and to suffer the consequences. As Frederick Schauer notes, "…the way in which and the extent to which, if at all, rules become a part of a decisional process is ultimately determined by the decision-maker alone"[348]. There are differences in the level of discretion included in the regulation, but in the end, the individual always makes the final decision[349].

Philosophy of law provides an explanation of the way legal norms guide behaviour. In the traditional understanding, legal norms guide behaviour by providing and communicating a general standard of conduct that makes it possible for the addressees to evaluate their behaviour in a specific situation without further direction[350]. However, as Raz points out, not even the recognition that standards provided by regulation serve as basis for evaluation for individuals when they make decisions on how to act clarifies the way the guiding function is performed[351]. Neither does it distinguish between the various modes in which the law can and does guide behaviour. According to Joseph Raz the only way in which laws can guide behaviour is by adding to

---

[348] Schauer, *Playing by the Rules*, p. 128, makes this point in his philosophical examination of rule-based decision-making, while analysing the means by which prescriptive weight is added to the process of decision-making and how much difference does it make.

[349] After all, calculated and negotiated non-compliance are common phenomena as pointed out by Terence Daintith in Law as Policy Instrument, p. 29, in relation to the economic sphere in his introductory essay to a larger collection of studies on the instruments used in the implementation of economic policy and the role of law as an instrument for implementation. This holds at least for the user perspective into the law explicated, e.g., by Hydén in *Rättsregler*, p. 22-23, as separated from the perspective of the judges and regulators more generally, the perspective used in legal education, whose conduct might be more constrained by the law and the rulings in law. The user of the law, i.e., the object of regulation, looks at the reality and her position in it, and concentrates on what the valid law means for her economic or social action. The focus is on the consequences that the law is expected to have on the practiced area. The law is a strategic variable that sometimes leaves more or less discretionary power. Basically, the law is seen a risk that has to be managed on the basis of some type of more or less conscious cost-benefit calculation.

[350] This has been explicated, e.g., by H.L.A. Hart in *The Concept of Law*, p. 124.

[351] Raz, On the Functions of Law, p. 281.

the natural consequences of actions the additional consequences provided by the law[352]. These additional consequences affect the reasons for or against the performance of the actions[353].

By applying this approach to regulation more widely we can see that regulation guides behaviour by communicating that consequences follow upon the performance of certain actions. The regulation makes the consideration of the consequences relevant to the desirability of the actions bringing them about. Basically the regulation provides additional reasons for performing or abstaining from certain actions that, at the same time, are incentives for the object of regulation to behave in a certain manner. When people make decisions to perform or abstain from certain action they now not only have to consider the natural consequences of their actions but also the additional consequences provided by the regulation.[354]

---

[352] Raz in On the Functions of Law, p. 281-283, provides this explanation as an adjustment and a clarification to the traditional legal theoretic view of how a legal norms guide behaviour.

[353] As pointed out by Redondo in *Reasons for Action and the Law*, p. 1, in studying on the various roles the notion of reason for action plays in jurisprudence, some of the most prominent contributions to the concerns in contemporary philosophy about the idea of law-imposed duty and its relevance for action, such as Hart, Raz and von Wright, have their starting point the notion of "reason for action". According to them, the concept of reason is a necessary element for understanding the relationship between norm and action. Note that the very idea of a reason for action has generated an enormous literature in philosophy as pointed out by Schauer in *Playing by the Rules*, p. 112, with extensive references. By no means can I drawn on this literature in depth. This is left for further studies.

[354] As Frederick Schauer, *Playing by the Rules*, p. 8, points out in a philosophical examination of rule-based decision-making, the force of rules in pressuring behaviour comes from the sanctions (positive or negative) that attach to following or violating rules, or from attitudes about rules held by the objects of regulation. Schauer concentrates not only on legal rules, but on regulative rules in general, be they used by formal institutions like churches and associations, or less formal social practices like morality and etiquette (Schauer, *Playing by the Rules*, p. 12).
By concentrating on the consequences attached to actions by regulation, I take a more restrictive approach than Schauer to the analysis of the ways in which regulation provides reasons for action. Not only punishment and reward affect

Even though the conception of regulation (especially the law) as an incentive for changing behaviour is used especially and most forcefully in law and economics (implicit prices), and in regulatory studies in relation to the instruments that provide economic incentives[355], it is applicable, and have for long been used, more widely. Regulation (and especially law) as an instrument of public or private policy is built in the conception of regulation causing an effect on the behaviour of its addressees; regulation can be used to achieve certain desired goals. When the effects are caused through the decisions and choices of an object of regulation, what regulation can do is to provide reasons for action or inaction (i.e., either negative or positive incentives)[356]. It is not only economic incentives that the law or regulation provides; they provide reasons for performing or abstaining from a certain action in general.

---

the reasons for action for Schauer, even though he does recognise their importance for most decision-makers; also other prudential considerations, like the role of rules as serving an important role in decisional simplification or the role of rules as providing reasons for action by virtue of the decision-maker's distrust of her own capacities with respect to some family of decisions, and moral considerations, like the arguments for taking a rule to be a reason for action stemming from agreement with and willingness to cooperate in the system from which the rule emanates (e.g., the solution of Prisoner's Dilemma or coordination problems), can influence individuals decision to treat an applicable rule as a reason action (Schauer, *Playing by the Rules*, p. 123-126). However, as Schauer notes in *Playing by the Rules*, p. 123, by extending the consideration above sanctions and rewards, he adopts the perspective of the objects of regulation. Since these considerations are not provided by the rules themselves, they are excluded from the normative analysis in order to be able to see how normative guidance communicates itself into the behaviour of the objects of regulation.

[355] According to this incentive approach, which is often studied as the opposite of traditional command-and-control regulation, the objects of regulation can be induced to behave in accordance with the public interest by the regulator imposing negative or positive taxes, or by deploying grants and subsidies from the public purse. (Baldwin and Cave, *Understanding Regulation*, p. 41-42)

[356] Note the similarity to the economic analysis of the law. The mechanisms are basically the same. Different consequences generate different reasons for action similar to different prices providing different incentives for behaviour. Only the consequences attached to action by regulation are widened.

Since I widen the use of the notion of reason for action to the regulation in general, I commit myself at this stage only to the least restrictive understanding of this notion. I accept that regulations can provide new reasons for action[357]. I say nothing about their capability to exclude other reasons for action. I do understand that the difference in whether rules in general and the law especial is conceived as providing special reasons for action (protected reasons for action in the sense made explicit by Joseph Raz[358]) or only as providing additional first-order reasons for action, is genuine and important for the theories that explain the central themes of legal theory[359]. However, since my interest lies simply in the basic assumption these theories hold, i.e., their starting point in that the law can influence behaviour only by providing reasons for action, I do not go deeper into the wide theoretical discussion of the role of reasons for action[360].

---

[357] The argument that rules often function as a reason for a person to act in some way is generally accepted in legal theory as pointed out by Mark Van Hoecke in *Law as Communication*, p. 76.

[358] Raz, *The Authority of Law*, p. 17-19, argues that mandatory rules generated by authorities are protected reasons for action, that is, a combination of first-order reasons for action that directly affect or refer to actions (i.e., claim that one should act in conformity with them), and exclusionary second-order reasons for not relying on considerations excluded by the rule. As Michael S. Moore explicates in Authority, Law, and Razian Reasons, p. 849, while discussing the central presupposition of Raz's theory of authority and law of the kinds of reasons for action that authoritative legal norms give, second-order reasons are reasons for acting for reasons, whereas first-order reasons are simply reasons for acting.

[359] In legal theory, as pointed out by Redondo in *Reasons for Action and the Law*, p. 97, the way the law affects behaviour and especially the notion of reason for action is subject to a debate that is the heart of the discussion of the three central ideas of legal theory, i.e., the normativity of legal provisions that regulate behaviour, acceptance as a condition for the existence of a legal system, and the justification of legal decisions.

[360] Similar to Schauer in *Playing by the Rules*, p. 112 footnote 1, I try to avoid taking unnecessary positions with respect to the nature of reasons for action.

Despite the huge variety in consequences attached to actions provided by regulation[361], two overall ways of how the additional consequences determined by regulation influence the decision-making can be recognised. When people consider the additional consequences provided by regulation while making decisions to perform or abstain from a certain action the consequences stipulated by regulation either externally constrain the set of available options or influence the intrinsic desirability of the actions. The reason for action can therefore be expressed in both of these forms as noted in the philosophy of law by Raz[362]. Thus, under the basic assumption that regulation operates through the decision-making processes of the addressees (actor always chooses whether or not to behave in a certain manner), two basic ways regulation can influence behaviour can be separated; either by *externally constraining* the set of available options or by *influencing the intrinsic predispositions*[363] of their targets.

---

[361] Raz in On the Functions of Law, p. 282, remarks that the legal consequences attached to actions only by law as such are immensely varied. The consequences attached to action by regulation can be, in addition to legal consequences such as liability for a breach of duty, criminal and administrative sanctions (e.g., deprivation of liberty or property, and fines), or the duty to perform a contract resulting from having concluded it, also plain rewards such as status, prestige or money, or even immaterial consequences such as social pressure, feelings of guilt and proud. These non-legal consequences may be imposed without invoking the legal process. Due to the huge variety of possible consequences and the amount of overlap no overall list of different consequences is useful.

[362] Raz, On the Functions of Law, p. 282.

[363] This is a term provided by an influential sociologist Amitai Etzioni, university professor and director of the Institute for Communitarian Policy Studies in George Washington University, in his effort to combine the different perspectives that law-and-economics and law-and-society studies have on social norms into a single approach called law-and-socioeconomics (Etzioni, Social Norms, p. 161). Etzioni uses the term to refer to the directions in which actors would channel their efforts if left to their own devices. These predispositions reflect a combination of people's biological urges and their cultural imprinting. Specified predispositions are often referred to as preferences (especially in economics). However, intrinsic predispositions include preferences, but encompass other concepts as well and can be less specified. According to Etzioni, the different understandings of social norms

This bears resemblance to the separation between external and internal social control processes in social control studies as explicated in sociology[364] and in social theory[365], to the separation of external sanctioning and internalisation in the enforcement of social norms[366] and to the two processes that the law has been seen to entail as an instrument of social change[367]. With the former reference is typically made to a social process whereby people conform to norms or rules because they are rewarded with status, prestige, money and freedom when they do adhere to them and are punished with the loss of them when they do not. This process is sometimes called coercive, external, or just social control. In the sociological studies of social norms, this is sometimes referred to as institutionalisation[368] or social enforcement.

---

(merely as a part of actors' environment or also affecting their intrinsic predispositions), is visible in the law-and-economics and law and socioeconomics type of studies (Idem.).

[364] See, e.g., Liska, Modelling the Relationships Between Macro Forms of Social Control, p. 39-40

[365] David Garland in *Punishment and Modern Society*, p. 252, points out that penality, a term he uses to refer to the complex of laws, processes, discourses, and institutions which are involved in the whole process of criminalizing and penalizing (idem. p. 10), acts as a regulatory social mechanism in two distinct respects: it regulates conduct directly through the physical medium of social action, but it also regulates meaning, thought, attitude – and hence conduct – through the expressive, symbolising function of penal practice that he calls 'signifacation'.

[366] This separation is used, e.g., by Horne, Sociological Perspectives on the Emergence of Norms, p. 4-5, in providing a short overview of the approaches of sociology to the enforcement of social norms in a recent collection critical essays studying social norms from a variety of disciplines such as sociology, economics, game theory and legal studies.

[367] Evan in Law as an Instrument of Social Change, p. 555, separates between two different but interrelated processes that the law entails as an instrument of social change. Institutionalisation of a pattern of behaviour means the establishment of a norm with provisions for its enforcement and internalisation of a pattern of behaviour means the incorporation of the value or values implicit in a law.

[368] According to Parsons et al. in Some Fundamental Categories of the Theory of Action, p. 20, note 26, institutionalisation of a norm means that the members of a group reward conformity to the norm and punish deviation

With the latter reference is made to a process where people adhere to social norms because they believe in them, feeling good, self-righteous, proud, when they do adhere to them and feeling bad, self-critical, and guilty when they do not. Conformity to a norm becomes a motive of its own[369]. This process is sometimes called socialisation or internalisation.

Even though this dichotomy has been made in relation to social norms and the law, as influence mechanisms these two can be generalised to concern all regulation. Postema points out from the perspective of the philosophy of the law the wider applicability of this dichotomy to regulation more generally in discussing the way the law influences the practical deliberation (reasoning) of a person and thus his decision to act, i.e., how the law influences behaviour: "First, something can influence practical deliberation externally by so shaping the environment within which action can be taken that only some alternatives appear feasible or reasonable. Second, something can structure deliberation internally by providing directives about what should or must or ought to be done. Whereas external devices put obstacles in the way of rational agents choosing certain alternatives, internal devices provide rational agents with reasons why they should choose certain actions from the range of available actions."[370]

This dichotomy has been picked up also in the discussion about the possibilities to regulate behaviour in cyberspace. Lawrence Lessig uses similar notions in his largely successful effort to shift the legal culture of cyberspace from essential immunity from control (especially government's control) into the possibilities of regulation[371]. Lessig develops his distinction between the subjective and the objective

---

from it (Opp, The Emergence and Effects of Social Norms, p. 777).

[369] Opp, The Emergence and Effects of Social Norms, p. 777.

[370] Postema, Positivism, I presume?, heading I.A. paragraph 3.

[371] Lessig, *Code and Other Laws of Cyberspace*, p. 237-238, especially endnote 8 of the Appendix. Lawrence Lessig, Professor of law at Stanford Law School and the founder of the school's Center for Internet and Society, is one of the most influential person behind the change of the legal culture of IT professionals towards the possibilities of regulation in cyberspace.

perspectives[372] further into a question of general applicability of the process of internalisation; not just a mechanism of social norms but also of other regulatory instruments: "Once internalized, norms no longer need to be enforced to have force; their force has moved inside, as it were, and continues within this subjective perspective. In my view, we should see each constraint functioning in the same way: we subjectively come to account for the constraint through a process of internalization"[373].

For example, technology can be altered to allow or restrict certain behaviours and thus affect externally. However, it does more. It can also be used to facilitate and encourage or discourage certain behaviours. A low fence may not keep a determined trespasser out

---

[372] Lessig in *Code and Other Laws of Cyberspace*, p. 237 separates between the *objective* (that of someone observing when a constraint is imposed) and the *subjective* (that of the person experiencing the constraint) perspectives in considering the time a specific regulation imposes its constrain (or other effect) on an individual. The objective perspective bears resemblance with that of the decision-makers and the subjective perspective with the point of view of the objects of regulation in the classification of perspectives into the study of regulation made in section 1.5 above. From the objective perspective there is a difference between regulatory tools that demand payment up front (e.g. technology and market forces via payment) and constrains that let you play and then pay ('law' and social norms). Even though this difference exists from the objective perspective, it necessarily does not do so from the subjective point of view; "a constraint may be *objectively* ex post, but experienced *subjectively* ex ante" (Lessig, *Code and Other Laws of Cyberspace*, p. 237, original italics). With this separation Lessig also makes an argument about the need for subjectivity in regulation: "Law and norms are more efficient the more subjective they are, but they need some minimal subjectivity to be effective at all. …this is not the case with architecture. Architecture can constrain without any subjectivity" (Lessig, *Code and Other Laws of Cyberspace*, p. 238-239). About the importance of this distinction for social norms, see Etzioni, Social Norms, p. 161-165.

[373] Lessig, *Code and Other Laws of Cyberspace*, p. 237-238, especially endnote 8 of the Appendix. As we already noted, the conception of regulation that Lessig uses has no requirement of intentionality or even actors that regulate. To Lessig regulation is not just an activity, it is also an outcome. His conception of regulation includes every meaning from the table of regulation provided by Black above (Table 1-1). However, this does not diminish the value of the argument of the mechanism of internalisation being applicable generally to different forms of regulation.

of a backyard, but it may deter trespassers by sending a message about the wishes of the property owner of the acceptable sphere of behaviour. The message that the external constraint on behaviour transmits might also be contradictory to the purpose of the constraint. For example, the purpose of bars or boards on windows is to prevent entry. However, the expressive meaning of those bars is "I don't trust people" or "I believe this is a high crime area"[374]. The point is that when technology is altered for regulatory purposes or another regulatory instrument is used, not only the purpose of the regulation should be kept in mind but the meaning it transmits as well.

## 3.2 As internal influence

Even thought it has been widely acknowledged that regulation can influence behaviour internally, the explication of the way regulation changes intrinsic predispositions has turned out to be problematic. For example, even though it is widely agreed in sociology that internalisation of norms has the potential to increase compliance, the contribution of internalisation is still unclear. Even sociology still has much to learn about the factors that contribute to the persistence of internalised notions of oughtness[375]. Amitai Etzioni provides one explanation while considering the ways in which compliance to social norms that relies largely on external factors can be converted into compliance that relies mainly on intrinsic forces[376].

---

[374] Levin, Student Note, p. 114.

[375] This argument has been made by Christine Horne, Assistant Professor in of Sociology at Brigham Young University who has doctoral decrees in sociology and in law, in Sociological Perspectives on the Emergence of Social Norms, p. 21, in an effort to develop a framework with which to organise thinking about norms in a recent collection critical essays studying social norms from a variety of disciplines such as sociology, economics, game theory and legal studies.

[376] Amitai Etzioni in Social Norms, p. 163-171, discusses the different approaches to social norms, and especially the adherence to social norms, in the fields of neoclassical economics (the foundation of law and economics) and sociolegal studies or studies in law and society.

Social norms equate to the informal social control of the community. Non-legal, non-market rules defined broadly under the rubric of social norms profoundly affect human behaviour. Social norms are a social practice by which activities (existing apart from the practice) are controlled. Generally speaking, when legal scholars refer to social norms they are referring to informal social rules that individuals adhere to because of an internalized sense of duty, because of a fear of external non-legal sanctions, or both[377]. Social forms, on the other hand, are a social practice by which activities are (wholly or partly) constituted[378]. An understanding of these rules, which mainly comes from sociology, makes it possible to harness them for regulatory purposes[379].

---

[377] This point has been made, e.g., by McAdams in The Origin, Development, and Regulation of Norms, paragraph 2. As Etzioni notes in Social Norms, p. 161-165, there is a separation between the external (the location of the factor that induces compliance is external to the person experiencing the constraint – the relevant society or group) understanding of social norms and the internal viewpoint (the factor that induces compliance lies within the individual person) among law and economics scholars. The former is represented, e.g., by Lessig in The New Chicago School, p. 662, and in The Regulation of Social Meaning, p. 1044, and by Ellickson in The Market for Social Norms, p. 3. The latter is represented most visibly by Cooter with his long line of research developing the theory of internalisation of social norms; people obey and enforce norms because they internalise its normative element in a way that it becomes part of their preferences. See, e.g., Cooter, Law and Unified Social Theory, p. 61-66.

[378] As Adler, Expressive Theories of Law, heading III "Why Expressive Theories are Unpersuasive: A General Argument", notes with reference to John R. Searle (1969) Speech Acts: An Essay in the Philosophy of Language, p. 33 and idem. (1995) The Construction of Social Reality, p. 27-29, in his criticism of expressive theories of law, this dichotomy is similar to the one between regulative rules that govern antecedently or independently existing behaviour and constitutive rules that create or define new forms of behaviour. The distinction between social meaning and social norms made by Lessig in The Regulation of Social Meaning, p. 957, is best understood as a distinction between social norms and social forms.

[379] Note that despite the importance of social norms to sociologists, there is little consensus between the diverse disciplines of sociology even about the whole notion of social norms (what they are, how they are enforced, and how they emerge) as Christine Horne points out in Sociological Perspectives on the Emergence of Social Norms, p. 3.

Following Etzioni, the specific process by which social norms influence peoples' intrinsic predispositions, thus affecting compliance with regulation, are internalisation and persuasion: preferences are initially set by internalisation and thereafter they are subject to persuasion[380]. In this conception, internalisation is an element of socialisation whereby the actor learns to follow the rules of behaviour in situations where there are impulses to transgress and no external surveillance or sanctions[381]. This is accomplished through such non-rational processes as identification with authority figures and affective attachments. With persuasion Etzioni refers to non-rational processes through which adult preferences are changed, such as identification with authority figures, generation of group enthusiasm through rituals, and relation of new forms of behaviour to values that the person already holds in high regard[382].

These elements of socialisation are applicable more widely than just in relation to social norms[383]. The cognitive-structural-stage theories of moral development have been utilised also in analysing the development of the reasoning of the individual in relation to legal, political and moral socialisation[384]. However, it is norms as guidelines

---

[380] Etzioni, Social Norms, p. 166-171.

[381] Etzioni, Social Norms, p. 167, takes this definition from the psychological research of Lawrence Kohlberg on the cognitive development (of their moral judgement as their mature intellectually) of all human beings through several stages of development. Reference is made to Kohlberg L (1968) Moral Development, in Sills D (ed.) International Encyclopedia of the Social Sciences.

[382] Etzioni, Social Norms, p. 169.

[383] In dictionary terms, socialisation is the process whereby an individual learns to adjust to a group (or society) and behave in a manner approved by the group (or society) ("socialization." *Encycloædia Britannica* from Encycloædia Britannica Online, http://search.eb.com/eb/article-9103216 [28.2.2006, requires authentication])

[384] As Tapp and Levine note in Legal Socialization, p. 122 and 126, in developing their theory of legal socialisation and the development of individual reasoning towards the law, the socialisation strategies, whether moral, legal or political, share common features regardless of their specific content or context. I.e., there is considerable overlap despite the differences in emphasis among

for action that are reported and transmitted in the process of socialisation in these fields[385]. As Etzioni points out, social norms are central to the internal influence of regulation: intrinsic predispositions are formed in part by social norms and thus can change over time as social norms are changed[386]. Even though social norms do not act in isolation, they are a major factor among the elements that that shape predispositions and are the basis of individual choices[387].

But how can these somewhat unconscious and non-rational processes be harnessed for regulatory purposes[388]? Is it possible to influence the process of socialisation with the sustainable and focused attempts involved in the process of regulation and into the direction desired the regulator?

This is currently being discussed especially in law and economics research under the rubric of "expressive theories of law"[389] and is the

---

the three areas in terms of socialising contexts, indicators of authority and ethical concerns. Note that the legal levels theory has substantially been influenced by the cognitive-structural-stage theories of moral developmental as introduced by Jean Piaget and advanced by Lawrence Kohlberg and builds also on jurisprudential frameworks of Lon L. Fuller and John Rawls (Tapp and Levine, Legal Socialization, p. 123-126).

[385] Hechter and Opp, What Have We Learned about the Emergence of Social Norms, p. 395. Fine in Enacting Norms, p. 157, speaks of norms as narrated when he refers to verbal exhortation as a way to achieve socialisation. Normative narratives can be institutionalised obligations, i.e., behavioural prescriptions accompanied with the possibility of invoking institutional punishments, or narratives about the world that surrounds the actors, stories that encourage or discourage forms of action.

[386] Etzioni, Social Norms, p. 166.

[387] Etzioni, Social Norms, p. 163.

[388] That these processes are somewhat unconscious and nonrational does not mean that the actors have no means to liberate themselves from the constitutive influences of social norms. They can become more aware of the forces that shape them, including social norms, and they can work with others to change these forces. What it means is that there are considerable limits to the actor's ability to do so as Etzioni emphasises in Social Norms, p. 170.

[389] I am not interested in the controversial expressive theories as such. Instead, I am interested only in the obvious, but often overlooked issue that the meaning of regulation has impact on behaviour as such. It should be noted, as

basic starting point of information law as understood at least in the Nordic tradition[390]. The law can influence behaviour 'expressively', by the message that it embodies. This can happen even independently of the other non-expressive influences such as deterrence (included in sanctions and their enforcement). Without going deeper into the possible specific mechanisms by which meanings affect behaviour, which are still disputed[391], this mode of influence has been recognised

---

Dharmapala and McAdams point out in The Condorcet Jury Theorem and the Expressive Function of the Law, p. 1., that the consequentialist claims being discussed here can be distinguished from the various non-consequentialist arguments that law has a moral value based on what it expresses, independent of its effects. For a discussion of such theories, see, e.g., Adler, Expressive Theories of Law, p. 1363-1502; Anderson and Pildes, Expressive Theories of Law, p. 1503-1576; Adler, Linguistic Meaning, Nonlinguistic 'Expression' and the Multiple Variants of Expressivism, p. 1577-1595.

[390] This has been explicated in the Nordic doctrine, e.g., by Tuomas Pöysti in *Tehokkuus, informaatio ja Eurooppalainen oikeusalue*, p. 133-135 with references to Ethan Katsh, *The Electronic Media and the Transformation of Law*, Oxford University Press, New York, 1991, p. 6 and to Peter Blume, *Ret Som Information*, in Ånd og Rett, p. 195-202, Oslo, 1997. In information law the guidance provided by law is based on the force of expression and information – that information and communication can constitute and alter social reality. However, this type of research has not clarified how the provision of information influences behaviour. The argument is more general. At the basic level the law is formed of the communication to the norm addressees, i.e., of information; law is a message and the communication of that message is at the core. Information, by the means of communication, conveys a message of the existence, meaning and objectives of the legal norm; it gives a standard for behaviour for the addressees and a set of expectations for the behaviour of the object of regulation. The legal system as a societal institution is seen to be based on the processing, management, and communication of information.

[391] For a short overview into some of the modelling done in law and economics of these mechanisms see Dharmapala and McAdams, The Condorcet Jury Theorem and the Expressive Function of the Law, p. 1-3. There is also preliminary empirical support for one of the models according to which law influences behaviour expressively, by creating a focal point around which individuals coordinate (McAdams and Nadler, A Third Model of Legal Compliance). As McAdams and Nadler note, their work implies neither a rejection of any external models nor any other expressive theories (McAdams and Nadler, A Third Model of Legal Compliance, p. 5)

widely[392]. To the degree instruments are justified and their enactment is somehow expressed (made public by providing information about them) by the regulator or an intermediary acting on its behalf, they also can influence behaviour expressively.

Robert Cooter has explained how the proclamation of a new obligation and the explanations given about the reasons of the enactment are related to the expressive effects of laws: "Promulgating a law often involves proclaiming a new obligation, describing the sanction attached to its violation, and explaining the reason for enacting it. … These three parts of a law relate especially (but not uniquely) to the three consequences of a law that I explained. Proclaiming a legal obligation gives people instructions on what to do, which especially promotes the coordination of behaviour. Attaching a sanction to an obligation especially deters its violation. Explaining the law ideally convinces citizens to follow it. In brief, the three aspects of promulgating a law especially aim at expression, deterrence, and internalization."[393] The way for regulation to shape intrinsic predispositions is thus to express the meaning of the norm (its content, the behaviour it expects from the addressees) and to provide justifications for the norm (which by means of convincing enhances internalisation). Accordingly to the decentring thesis also other regulators than just the government can use these expressive and justificatory (explanatory) means[394].

Note that not only the expressive function of law shape intrinsic predispositions. Interaction between law and preferences can occur

---

[392] This influence mechanism of expressions is not denied even by the critics of expressive theories such as Adler, Expressive Theories of Law, heading "Conclusions". Adler notes that that the issue is not only uncontroversial, but banal.

[393] Cooter, Do good laws make good citizens?, p. 18-19. Also Black in Enrolling Actors in Regulatory Systems, p. 69, has recognised the issue that the process in which instruments (e.g. written norms) are produced and the manner in which they are expressed can both affect behaviour.

[394] The possibility for also other regulators than just the government to influence social meaning has been recognised by Lessig in The Regulation of Social Meaning, p. 957.

also via a payoff-based evolutionary model as suggested and preliminary tested by Oren Bar-Gill and Chaim Fershtman[395]. However, due to the overlap of both of these models, which is also acknowledged by the writers[396], and the problems in mobilising the evolutionary model in the intentional action of regulation, this line of inquiry is not taken further in this study. It is important to note, however, that not all changes in preferences initiated by legal or more widely regulatory rules can be explained only on expressive or symbolic grounds.

What is important for our purposes is that expressive effect of regulation (intrinsic predispositions shaping) is mediated primarily by social norms or social forms; the meaning of a regulation has causal effects on social norms or forms and it can be used to shape them[397]. This means that when regulation attempts to change intrinsic predispositions, the effects are mediated by social norms or forms. The influence mechanism is not direct. Instead, the information a regulation provides of itself (the justifications and the mere expression) affects social norms or social forms that then influence behaviour.

Changes in intrinsic predispositions do not necessarily, or even primarily, happen via rational models of decision-making. As Etzioni points out in noting that while socioeconomics and studies in law and society assume that intrinsic predispositions are formed in part by social norms and thus can change over time as social norms are

---

[395] Bar-Gill and Fershtman, Law and Preferences, p. 331-352. Acknowledging that lawyers, philosophers and psychologists have for long recognised the role of law in shaping norms and preferences (idem. p. 332), Oren Bar-Gill and Chaim Fershtman for their part try to expand the boundaries of law-and-economics by introducing an alternative (not exclusive) mechanism beside the expressive or symbolic impact of the law through which the law influences the formation of preferences as part of the analysis.

[396] Bar-Gill and Fershtman, Law and Preferences, p. 333.

[397] See Adler, Expressive Theories of Law, heading "Introduction", paragraphs 18 and 27-28, last paragraph of heading II "Expressive Theories: The Main Examples", and paragraphs 3-7 of the heading "Conclusion", citing the works of most recent expressive theorists Cass R. Sunstein, Robert Cooter and Dan Kahan.

changed, they also assume that these changes can take place through non-rational processes[398]. As pointed out by Janet Weiss in her analysis of public information as a governance tool, information may affect both the explicit and implicit cognition, i.e., people may still respond to information in the world around them and connect it to their subsequent behaviour even though they might not be making deliberate, conscious calculations about what they do[399].

As Etzioni points out, knowledge does not necessarily affect intrinsic predispositions and increased knowledge together with the developed capabilities for reasoning in terms of high level principles (the most advanced type of individual reasoning in relation to morality and law) does not necessarily mean that actual behaviour is changed: "to know the good cannot be equated with doing the good"[400]. Increased information does not necessarily lead to changes in behavioural patterns. In relation to the values implicit in a regulation, i.e., the norms to be internalised, increased knowledge can lead to opposite direction. As research on the support for the legal system in the United States indicates, support for the legal system declines as knowledge of it increases, i.e., knowledge of the legal system and support for it may be inversely related[401].

As such this basic understanding of the intrinsic predisposition shaping mechanism is sufficient for the purposes of this study. To note that the publishing (proclaiming) a norm, the expression of its meaning, and the justifications given all influence behaviour, gives the possibility analyse the cause it can have on the regulatory capability of different instrument. It makes it possible to see that the meaning

---

[398] Etzioni, Social Norms, p. 166.

[399] Weiss, Public Information, p. 218-219 and 227-228.

[400] Etzioni, Social Norms, p. 168-169.

[401] Sarat, Support for the Legal System, p. 181. However, the argument that support for the legal system declines as knowledge of its actual operation increases holds only for the satisfaction with the legal system as a part of support for it. Willingness to comply does not decline in similar manner. In cases where dissatisfaction persists over time, the willingness to comply may also be eroded.

a regulation conveys might be different from, or even contradictory to, the objectives of the regulator or the substance of the norm. However, the concept of intrinsic predisposition shaping mechanism is still too underdeveloped theoretically to be used in any deeper analysis of the influences mechanism of regulatory instruments[402]. Its mechanism are simply too complex and unknown.

It needs to be acknowledged that these methods to shape intrinsic predispositions can be used in relation to all regulation. Information *on* instruments (understood as a metainstrument) can be used to affect behaviour in relation to all of the different types of regulatory instruments[403], including the informative and pedagogic instruments themselves (information *as* a policy instrument)[404]. For example, when a new informative material has been developed, the regulator has to

---

[402] As Davind Garland, *Punishment and Modern Society*, p. 251, notes in a discussion of the major social theories of punishment, the analysis of penal practice as a form of social signification (i.e., the expressive symbolizing function of penal practice) inevitably involves interpretative statements backed by illustrative examples rather than by solid evidence and that theoretical arguments outrun the available data. This is still true in relation to the intrinsic predisposition shaping mechanism of regulation, despite the efforts of many disciplines of which the behavioural law-and-economics has been most active recently. Oren Bar-Gill and Chaim Fershtman, Law and Preferences, p. 346-347, also acknowledge the controversial nature of preference formation and the role of law in it.

[403] There is a huge variety of tools that different regulators can use to strive for the desired objective by attempting to alter the behaviour of others. A classification of instruments is provided below.

[404] It is illustrative to separate between information *as* a policy instrument and information *on* policy instruments. Information (together with education) *as* a policy instrument is a distinctive and separate category as such. The tools used in this category include, for example, the transfer of knowledge, communication or reasoned argument, persuasion, advice, moral appeals etc. However, in addition to information being a type of instrument in its own right, it also is metainstrument in the sense that it is used to disseminate knowledge of the existence, meaning, and availability of other regulatory instruments. Vedung, Policy Instruments, p. 48; Vedung, *Public Policy and Program Evaluation*, p. 133-134. This is a clarification to the central notion made in the sociology of law about the centrality of information about the content of laws to the guiding function that a law can have, of which see, e.g., Mathiesen, *Rätten i samhället*, p. 62-67.

decide whether it should disseminate the material immediately to the target group or whether the potential users of the information should be informed about the availability of the material and its importance, and that prospective addressees should order it.[405]

For analytical purposes the fact that regulation may influence behaviour by the message is embodies and by its communication, i.e., as information *on* policy instruments (information as a metainstrument) because it shapes intrinsic predispositions, needs to be kept separate from the possibility that it may influence behaviour also externally. Information can be provided merely in the purpose of giving information about the specific instrument, its existence, and the behaviour it expects from the addressees (the standard of conduct provided); basically the way it operates as an external constraint[406]. The information affects the considerations of the external circumstances, the costs and benefits included. However, the mechanisms considered here by which information as a metainstrument affects behaviour is when information on instruments is provided in the purpose of appealing to actors subconscious motives, ranging from guilt to sexual desire, and thus to shape intrinsic predispositions.

A good example is advertising. Advertising is not strictly informational and thus not only affects behaviour by affecting considerations of costs and benefits (and, as a rule, do not shape preferences; i.e., advertising is considered to affect as an external

---

[405] Vedung, Policy Instruments, p. 48-49; Vedung, *Public Policy and Program Evaluation*, p. 134.

[406] This is the traditional case of horizontal packaging, i.e., a process identified in policy studies, where two of more instruments are directed simultaneously at the same target. See, e.g., Bemelmans-Videc and Vedung, Conclusions, in *Carrots, Sticks* & *Sermons*, p. 262. Compare to Vedung, Policy Instruments, p. 48, in the same book, who refers to this practice as vertical packaging. However, the packaging is horizontal because both types of instruments (information as a metainstrument and the instrument on which information is provided of) are directed simultaneously at the same targets and one is not used to entice some implementing actor to employ to other toward the objects of regulation, as with vertical packaging. More about vertical packaging see infra.

mechanism) but also a means of subconsciously affecting people's preferences and thus a form of persuasion in the terms of Etzioni[407].

## 3.3 As external constraint

As pointed out by Etzioni in relation to social norms, as external constraints social norms and regulation in general either increase or decrease the availability of choices, change the cost or benefits of decision, change the resources the actor draws on, or physically change the environment where the decisions are made in a way that restricts or allows certain behavioural patterns[408]. When regulation affects behaviour as an external constraint it is akin to the changes that take place outside the actor, which the actor includes (perhaps even unconsciously) in her calculations and choices.

This does not provide much help in the use of the external constraining mechanism as a heuristic device in analysing the ways specific regulatory instruments affect behaviour. The influence mechanism remains much too unspecified. Since there does not seem be specific influence mechanisms or a set of mechanisms that apply to all regulation as an external constraint, as is the case with the intrinsic predisposition changing mechanisms, a way to deepen the understanding is needed. One possible path is to look at the types of regulatory instruments[409].

---

[407] Etzioni, Social Norms, p. 169-170.

[408] Etzioni, Social Norms, p. 161-162. Tala, *Lakien vaikutukset*, p. 311, points out similar influence mechanisms in relation to law. When regulation and especially the legal variation of it (the law) is studied, the concentration typically is on the external constraint aspect. Resent studies on social norms have picked up the issue of how law and other instruments (especially technology) can change social norms. However, many of the studies have still treated social norms as external factors as pointed out, among others, by Etzioni in Social Norms, p. 161-165.

[409] The different types of regulatory tools, made explicit by classifying them on different basis, are what can also be called regulatory strategies. Thus, strategies are separated from the actual tools or instruments that the regulator uses in

There is a wide array of instruments that can be used by different regulators in attempting to alter the behaviour of others; the different instruments may be used by the state, the community, associations, networks, or individual actors (including firms) with the obvious exception within constitutional democracy that the use of force and imprisonment are confined to the state[410]. Actually there is not even a consensus on the number of tools that exist for governmental use[411].

Due to this huge variety the need to differentiate the issue for analytical purposes has been recognised. As pointed out by Evert Vedung, there are two fundamental approaches to the classification of instruments: the maximalist and the minimalist[412]. The *maximalist* approach makes long lists of possible instruments, but little effort is made to arrange them into smaller or larger groups. Since the maximalist approach just lists all the possible types without any particular ordering, it does not provide a taxonomy that might be used in considering the connection between the instrument and its effects. The *minimalist* approach searches for few basic types under which all specific kinds of individual policy instruments can be categorised. According to this approach there are, in principle, few basic types

---

specific situations. Tools are concrete and specified operational forms of attempting to influence behaviour. They are a specific way of making the strategy concrete in a specific situation. Strategy is more general; it concerns the regulators decision of the basic type of action it adopts and the general form of the regulation. Somewhat similar use of the concepts is in Bemelmans-Videc, Introduction: Policy Instruments Choice and Evaluation, p. 4 and in Baldwin and Cave, *Understanding Regulation*, p. 34-35. Tala uses the terms a bit differently in separating also between regulatory strategies and different types of tools in *Lakien vaikutukset*, p. 157.

[410] This is the approach taken both in relation to State's capacity to create and control 'self-regulation' as well as the non-state centred regulators possibilities to use different instruments already at the mid 1990´s by Daintith in Law as Policy Instrument, p. 32-33, and acknowledged also by Christopher Hood in *The Tools of Government*, p. 120-121. This is repeated as a normative argument in Black, Decentring Regulation, p. 139.

[411] Salamon, The New Governance and the Tools of Public Action, p. 21; Tala, *Lakien vaikutukset*, p. 209-210.

[412] Vedung, Policy Instruments, p. 22.

of tools that the regulator can use, even though they can be combined and changed for different contexts.

In the minimalist approach followed here the instruments are categorised into ideal types for analytical purposes. These ideal types of instruments are only heuristic tools and not the aims of the investigation. They do not represent the reality of what types of tools regulators use. Instead, they are tools used in order to begin a process of interpretation and evaluation of the tools used. The ideal instrument types are mental constructs (not depictions of reality) to be used for the analysis of the effects of the diverse set of instruments that can be used in practice.[413]

Different classifications of regulatory instruments have been used to collect information on the ways different types of instruments affect and what kinds of effects they generate[414]. The classification seems to underlie a conception of the connection between the instruments and their effects. By looking into types of regulatory tools and strategies the relationship between the type of regulation and the effects can be more easily seen; the classification of regulatory tools makes it easier to compare the different instruments. The different classifications of regulatory tools help to outline and to illustrate the connection between the instrument and its effects.[415]

---

[413] This resembles the Weberian ideal-type methodology as depicted by Weber in *The Methodology of the Social Sciences*, p. 99-100. It is important to note that the classifications do not reflect the interactions and the overlapping of regulatory instruments. After all, they play several roles and influence behaviour in a variety of ways. Thus specific regulatory instruments do not as such fit simply to just one ideal regulatory instrument type. They might be categorised into several models or placed somewhere in the between.

[414] As Tala points out in *Lakien vaikutukset*, p. 156 and 343-344, in relation to the legal instruments, the information collection has not been done systematically and especially not always by using the same classification. This is why the policy analysis and evaluation types of studies are useful in the consideration of the influence mechanisms of specific types of regulation. That is, to the degree they can be considered to have induced the mechanisms and effects of different instruments.

[415] As Tala in *Lakien vaikutukset*, p. 156, 179 and 210-211, notes, the classifications of regulatory instruments do not provide much information

There are a variety of different classifications of regulatory instruments; even of those that are available to the state actors[416]. Instruments have been identified and classified from a variety of perspectives and for several purposes and these classifications separate from one another on the basis of their explicatory force in relation to the connection between the instruments and their intended effects. Most basic classifications are made in order to see the alternative (to statutory or case law) policy tools or regulatory strategies of government or that of public action more generally. These classifications provide ideas and holds for the analysis of the content of regulation and their intended effects by making roughly explicit the normative guidance mechanism involved.

By analysing several classifications the effort in the following presentation is to illustrate the means by which different regulatory instruments provide reasons for action (i.e., how they guide behaviour, the way the effects are assumed to be generated at the normative level). Such an analysis allows the use of the different analytical forces in the classifications. A short presentation of the different classifications

---

about the way effects of regulation are generated or the relationship between the instrument and its effects, despite them underlying a conception of the connection between the instruments and their effects. The classifications do note tell much about the actual or intended effects, i.e., about their social functions. But they do tell about the normative guiding function; the different ways regulation can guide behaviour are visible. This is why they are relevant in the analysis of the influence mechanisms and need to be explicated in here. Cf. Tala, *Lakien vaikutukset*, p. 150, 208, 212-213 and 342.

[416] This has been pointed out in policy studies by Vedung in Policy Instruments, p. 22, and has been emphasised also by the OECD Working Group on Regulatory Management and Reform in *Regulatory Policies in OECD Countries*, p. 52. For examples of different classifications of tools of public action see, e.g., Tala, *Lakien vaikutukset*, p. 151-178; Baldwin and Cave, *Understanding Regulation*, chapter 4; Ogus, *Regulation*, parts III and IV; Salamon, The New Governance and the Tools of Public Action, p. 19-37; Breyer, *Regulation and its Reform*, chapter 8; Gunningham and Grabosky, *Smart Regulation*, chapter 2. There is even an overview of the classifications of the tools of government provided in König K and Dose N (1989) Klassifizierungsansätze staatlicher Handlungsformen, Speyerer Forschungsberichte 83, referred to in Tala, *Lakien vaikutukset*, p. 153.

together with their capability in linking the instruments with their intended effects and the way these classifications combine is in order.

### 3.3.1 Different classifications with a common background assumption

Classifications have been made, as noted in policy studies about the tools of public action by Salamon from the perspective of political science, e.g., on the basis of the degree of coerciveness, degree of directness, degree of visibility in the budgeting and policy review process, and extent to which a tool utilises existing administrative structures to produce its effects (automacity)[417]. The resources required for the operation of the tool[418] and the way the instruments affect their objects or the types of behaviours the instruments seek to modify have also been used[419].

1. One basic general classification of alternative policy tools from the point of view of the state, by far the most common[420], is made in the policy studies on the basis of the *degree of coercion* or the extent of authoritative force that each instrument involves[421]. This classifier is used, as has been pointed out by a Working Group on Regulatory Management and Reform of the OECD, especially in the economic

---

[417] Salamon, The New Governance and the Tools of Public Action, p. 24-37.

[418] This classificatory principle is used in the studies of regulation by Hood in *The Tools of Government*, p. 4-7.

[419] These classificatory principles have been used in sociologically minded legal studies by Eckhoff in *Statens Styrningsmuligheter*, p. 29-45, and in political science and especially studies on policy tools by Schneider and Ingram in Behavioural Assumptions of Policy Tools, p. 510-530.

[420] This is according to Salamon, The New Governance and the Tools of Public Action, p. 25 and 27.

[421] As Vedung, Policy Instruments, p. 34-35, points out, this is the degree of constraint it sets for its objects or power the regulator has invested in the attempt to regulate; the extent to which a tool restricts individual behaviour as opposed to merely encouraging or discouraging it.

studies because it measures the degree of intervention on the "free market" (the extent to which a tool involves a deviation from the reliance on the market as a mechanism to allocate resources and settle social roles)[422]. In a basic form this classification involves three types of tools: regulations, economics means and information (in popular terms, carrots, sticks and sermons)[423].

*Regulations* (sticks) are measures undertaken by governmental units to influence people by means of formulated rules and directives which mandate receivers to act in accordance with what is ordered in them; the threat of negative sanctions might be involved but is not a defining feature[424]. The authoritative relationship between the regulator and the objects of regulation, in the sense that they are obligated to act in the commanded way, defines regulation and separates it from other instruments. Thus, the term regulation is used in the strictest of the earlier mentioned three definitions; as agency regulation, ministerial decrees or detailed parliamentary laws.

*Economic instruments* (carrots) involve either the handing out or the taking away of material resources, in-kind as well as in cash. Their intention is to make it cheaper or more expensive in terms of money, time, effort, and other valuables to pursue certain actions. What

---

[422] OECD, *Regulatory Policies in OECD Countries*, p. 52 and annex 2. This has been pointed out also by Salamon, The New Governance and the Tools of Public Action, p. 25.

[423] Vedung, Policy Instruments, p. 29-34; Bemelmans-Videc and Vedung, Conclusions: Policy Instruments, Types, Packages, Choices and Evaluation, p. 250-257. These instruments have remained the same even in some of the decentred understanding of regulatory strategies (regulation as indirect intervention in the self-regulation of social and market actors). However, the ways they are used have changed, especially in relation to legal rules; there is a move from regulatory law (that sets substantive standards) to reflexive, procedural or post-regulatory law (that sets procedures. This has been emphasised by Black in Decentring Regulation, p. 126 with reference to Bruijin J and Heuvelhof E (1991) Policy Instruments for Steering Autopoietic Actors, in Veld et al (eds.) *Autopoiesis and Configuration Theory: new Approaches to Societal Steering*, Dordrecht, p. 161.

[424] Vedung, Policy Instruments, p. 31-32. Similarly in relation to legal norms, e.g., in Aarnio, *Laintulkinnan teoria*, p. 64.

separates economic instruments from regulation is a fact that the objects of regulation are not obligated to take the measures involved; they are just encouraged or discouraged to do so. The regulated behaviour is neither prescribed nor prohibited, instead, it is made more or less expensive or easier (or more difficult) by the adding or deprival of material resources[425].

*Information* (sermons, also referred to as moral suasion, exhortation, or public communication) covers attempts at influencing people through the transfer of knowledge, the communication of reasoned argument, and persuasion. No more than the plain transfer of knowledge or persuasive reasoning is offered to influence people to do what the government deems desirable; no government obligation or coercion is involved. This absence of obligation separates information from regulation and the lack of material resources handed or taken away separates it from economic instruments.[426]

2. There also is a somewhat similar classification, made on the basis of the *resources* available to the public authority[427]. In this scheme the regulatory instruments that can be used to affect the objects of regulation correspond to categories of government resources. The resources and the subsequent instruments are conventionally classified into four basic types. (1) *Nodality* refers to the central position of government in society's information flows and the consequent information resources. The instruments exploit the ability to use the information resource for different purposes, e.g., giving advice and information or to persuade. (2) *Treasures* are the financial and other material resources available to the state, in-cash or in-kind. Instruments use this wealth or money (e.g., the ability to raise tax finance and requisition of other resources such as land or the time

---

[425] It ought to be stressed that in this taxonomy economic instruments include non-monetary as well as monetary material resources. According to Vedung "a bumb in the road to prevent motorist from speeding is an economic instrument just as a tax levied on gasoline is" (Vedung, Policy Instruments, p. 32-33).

[426] Vedung, Policy Instruments, p. 33-34.

[427] Hood, *The Tools of Government*, p. 4-7 and 21-91.

of conscripted personnel). (3) *Authority* refers to the monopoly to give binding legal rules and the ability to compel compliance. Basic command and control instrument such as laws or regulations belong to this group. (4) *Organisation* refers to the human labour, buildings and devices and other physical assets the government possess and is able to employ. The capability to set up and administer basic bureaucratic organisations and employing them in direct action form the class of instruments involved.[428]

Despite the different basis of classification, these taxonomies are similar to a high degree. As Vedung also notes, nodality, treasure and authority seem quite similar to the categories of information, economic means, and regulation in the former threefold scheme[429]. Also the coerciveness argument is applicable to this classification. The instruments in this classification can be scaled on the basis of the degree of restriction they pose for their objects; starting from the use of informational resources and ending in the use of force by the organisation of the public authority[430].

The biggest difference is related to the organisational resource and the related direct government instruments (e.g., direct provision of goods and services). The former coerciveness-based taxonomy adopts a view, according to which organisation is a prerequisite for the application of instruments, not an instrument in itself[431]. Organisation is necessary for the provision of regulatory, economic and informative instruments; instruments cannot be applied if there is no government organisation. Even though this is a strict state centred understanding, it is applicable more widely. When a regulator uses its own

---

[428] In addition to these effectors (means by which the public authority can influence its external environment) also detectors are used to collect information required for the operation of public authority. Hood, *The Tools of Government*, p. 3. Note that Christopher Hood uses the same resource based classification also for the detectors.

[429] Vedung, Policy Instruments, p. 38.

[430] Hood, *The Tools of Government*, p. 7. See also Tala, *Lakien vaikutukset*, p. 153-154.

[431] Vedung, Policy Instruments, p. 37-38.

organisation to act directly with its own resources it is just making a choice in the mode of implementation of the instrument (e.g., provision of material resources directly instead of using subsidy). Thus, an employance or an implementation strategy should be separated analytically from policy instruments.

> Several other classifications of resources and the subsequent instruments have been made. For example, Baldwin and Cave in their basic textbook on regulation separate following basic capacities and resources that governments possess and which it can use to influence industrial, economic or social activity: the ability *to command* ( legal authority and the command of law is used to pursue policy objectives), to *deploy of wealth* (contracts, grants, loans, subsidies, or other incentives are used to influence conduct), to *harness markets* (governments channel competitive forces to particular ends), *to inform* (information is deployed strategically, e.g., so as to empower consumers), *to act directly* (state takes physical action itself), and *to confer protected rights* (rights and liability rules are structured and allocated so as to create desired incentives and constraints)[432].

What these classifications of resources do not take into consideration, and what is essential in the decentred understanding of regulation, is that these resources are dispersed among a range of actors; they are not confined to the state[433]. When the range of possible regulators is expanded by removing the state from the centre, the resources become more difficult to identify. Basically any assessment of what resources are necessary for regulation is likely to vary at the detailed level with the particular context[434]. However, Black has identified six key resources that are critical to the performance of at least one or more of the cybernetic regulatory functions (standard

---

[432] Baldwin and Cave, *Understanding Regulation*, p. 34-35.

[433] This has been pointed out by Daintith, Law as Policy Instrument, p. 33, in his introductory essay to a larger collection of studies on the instruments used in the implementation of economic policy and the role of law as an instrument for implementation. This is central to Black's decentring thesis as she makes explicit in Decentring Regulation, p. 139.

[434] Black, Enrolling Actors in Regulatory Systems, p. 73.

setting, information gathering and behaviour modification): information, expertise, financial and economic resources, authority and legitimacy, strategic position and organisational capacity[435]. But these resources are so dispersed and interrelated in a way that it is no longer possible to identify types of tools on the basis of them. The resources in each context with its own variations become more related to actors than to specific instruments. There might be a tendency to use certain instrument by a regulator that possesses certain specific resources (e.g., informational), but all of the different instrument can be used by every regulator.

3. The classification made by Eckhoff in sociologically biased legal scholarship on the basis of how the different types of tools influence their objects bears resembles with the preceding taxonomies: normative, economic, pedagogic and physical[436]. The *economic* instruments in this classification refer to those mechanisms that are used to affect the individual's considerations of what kind of behaviour is economically rational. Included are tools like taxing, or regulation by contract. The *normative* tools are basic command and control. Injunctions, permissions, indemnity rules etc. belong to this group. *Pedagogic* tools contain information and attempt to affect the attitudes of individuals. The *physical* tools are those that facilitate or hamper certain operations in the actor's environment. Examples are buildings, devices, and other physical constructions[437].

The difference to former coerciveness-based and the resource-based classifications is that whereas they include both in-cash and in-kind material resources into the economic/treasures subcategory, the influence mechanism –based taxonomy separates physical instruments into a separate category. Since the instruments seem

---

[435] Idem.

[436] Eckhoff, *Statens Styrningsmuligheter*, p. 29-45. See also Matthiessen, *Rätten i Samhället*, p. 89 where this classification made by Eckhoff has been used.

[437] The classification made in 1983 does not yet contain the changing of the logical environment (the virtual space) with tools like software. However, it could be included without further problems.

different and affect in distinctive ways, as argued below, this separation is useful.

Another classification made in policy studies on the basis of how instruments influence their objects is provided by Mayntz[438]. In her classification *regulative norms*, such as prohibitions supported by the threat of sanctions, rules of market entry, conditional permissions, produce direct effects. They shape behaviour according to a norm. *Financial transfers and incentives*, such as grants, subventions and welfare payments, instead work more indirectly. *Procedural regulation* is also indirect in its influence. It sets rules for the cooperation and/or resolution of conflicts between two or more parties without trying to control the outcome of these processes directly[439]. Norms establishing decision and conflict resolution procedures for private parties are mentioned as examples of procedural regulation. *Public provision*, such as technical and personal services and production of goods, is also a separate type of instrument in this classification. The final type of instrument is *persuasion*. It involves campaigns to inform and to exhort.

4. A classification on the basis of how the objects of regulation are expected to behave, and at the same time, how their behaviour (or lack of action) can be influenced, has been provided in political science and especially studies on policy tools by Ingram and Schneier[440].

If the target population does not believe that the regulation directs them or authorise them to take action, then tools providing *authority* can be used. They are simple statements backed by the legitimate authority of government that grant permission, prohibit, or require action under designated circumstances. If the objects of regulation lack incentives or capacities (e.g., information, skills, material resources) to take actions needed, then tools providing incentives or

---

[438] Mayntz, The Conditions of Effective Public Policy, p. 127-129.

[439] Mayntz, The Conditions of Effective Public Policy, p. 139.

[440] Schneider and Ingram, Behavioural Assumptions of Policy Tools, 1990, p. 514-522.

capacity can be used. *Incentive* instruments include tools like inducements (positive payoffs that encourage behaviour), charges (monetary charges for those who do not meet the standards involved or want to exceed their quota), sanctions, and the use of force in its direct physical action sense. *Capacity* tools provide information, training, education, and resources to enable decision making.

If the objects of regulation disagree with the values implicit with the instruments or their goals, then instruments using *symbolic or hortatory* proclamations to influence perceptions of values can be used. An example is the use of persuasive communications that seek to change perceptions about the regulated behaviour through appeals to intangible values (e.g., justice, fairness, equality) or through the use of images, symbols and labels. If the regulated situation involves, in turn, such high levels of uncertainty that the nature of the problem is not known, and it is unclear what people should do, then the promotion of *learning* to reduce uncertainty can be used. Included are, for example, participatory tools such as hearings, advisory boards, or citizen panels, and mediation and arbitration programs.

In principle, the basis of classification again is different in this taxonomy. However, even the classification on the basis of coerciveness includes similar considerations of how the instrument types affect the behaviour of their intended objects as the taxonomies of Eckhoff, Mayntz, and Ingram and Schneier. As noted earlier, the regulatory tools in the scheme of Vedung imply an influence mechanism of expected compliance which bears resemblance to normative and authority tools: "the governee is obligated to do what the governor tells her to do"[441]. Similar to pedagogic or certain capacity tools, the informative instruments attempt to persuade and affect the attitudes of individuals. In the case of economic tools the influence mechanism concentrates on making the constrained behaviour more or less expensive/difficult by the adding or deprival of material resources similar to incentive tools.

---

[441] Vedung, Policy Instruments, p. 31. See also Tala, *Lakien vaikutukset*, p. 292.

### 3.3.2 Attaching specific external influence mechanisms to instrument types

Acknowledging that every instrument can influence behaviour by shaping intrinsic predisposition (mediated mainly by social norms or forms) and associating it in relation to every instrument category, specific external and normative influence mechanisms can be roughly attached to every regulatory strategy on the basis of these different classifications. This is not to suggest that the instruments absolutely cannot influence or be used to alter behaviour in other ways. The statement is that the types of regulatory instruments seem to differ in their ways of influencing behaviour and that this forms the main basis for the categorisation.

As a combination of the several classifications presented above, the main types of regulatory tools are roughly classified in the following way in this study: Prescriptive rules (legal or non-legal), market-harnessing or economic, informative or pedagogic (social norms harnessing), technology-harnessing, and process-based. Note that this comes close to a general classification suggested by Black[442]. Her general taxonomy involves five main types of regulatory tools: written norms (legal and non-legal) and accompanying sanctions, economic- or market-based instruments, social norms and accompanying sanctions, technologies, and processes[443].

---

[442] Black, Mapping the Contours of Contemporary Financial Services Regulation, p. 3-7.

[443] Black, Mapping the Contours of Contemporary Financial Services Regulation, p. 3-4. Even though Black makes her classification in the connection of trying to show the partial inadequacy of tools-based analysis of decentred regulation and to suggest a new method of "enrolment analysis", it is worth noting especially because it bears resemblance to the analytical frame for studying the potential for alternative (to law, i.e., regulation in the strict sense or even as duty-imposing legal norms) forms of control in cyberspace popularised by Lessig in *Code and Other Laws of Cyberspace*, chapter 7.
Note the difference in the reason for the use of the concept 'regulation'. Lessig wishes to see the possible forms of control, which leads him to adopt a wide-ranging conception of regulation, covering every question from social and

> The names Black uses to refer to the types of instruments are rather surprising, especially in the case of markets, social norms, and technologies. As already noted above she has ruled them quite correctly outside the definition of regulation as an intentional process[444]. Then why are they classified as types of regulatory tools in here?
> Even though the patterns of interaction of rational actors (market forces), social norms or cultural world views, or technologies are not in themselves considered as regulation, it does not mean that they are irrelevant in the process of regulation. On the contrary, as Black also notices, they are influential in how regulatory systems operate, and the regulatory systems often seek to harness these or alter them for their own purposes[445]. In short, they are something constraining (setting the limits for) behaviour (including regulatory behaviour) and are harnessed in the processes of regulation and for its purposes. When they are harnessed into the process of regulation – regulators intentionally use them to modify behaviour – then they become regulation. Thus, they become part of the regulatory toolkit available to the regulators, together with legal and other prescribed norms. This is also the approach of Black and becomes evident in her analysis of the different types of instruments.

The category of *prescriptive rules*, incorporates 'regulations' or sticks from one of the basic forms of coerciveness-based taxonomies

---

political sciences and thus weakening the analytical resolution of the concept. For Lessig also markets, social forces and technologies 'regulate'; there is no requirement of intentionality. See, e.g., Lessig, The New Chicago School, p. 662. Black, instead, concentrates on the intentional attempts to regulate. By defining regulation as an intentional, systematic attempt at to alter the behaviour of others the categories of market forces, social forces and technologies are excluded. They do not regulate because there are no recognisable 'actors' having intentions to affect behaviour or, as with market forces, the outcome of the interactions of actors is not intentional in the sense of affecting behaviour for some defined purpose or producing broadly identified outcomes. The patterns of interaction of rational actors, or social norms or cultural world views, or technologies are not considered to be regulation as such. See Black, Critical Reflections on Regulation, p. 16 and 20.

[444] See above heading 1.4.

[445] Black, Critical Reflections on Regulation, p. 20.

proposed by Vedung[446], both authority and parts of organisational instruments from the resources-based taxonomy (those that concentrate on the enforcement) of Hood, the normative tools from Eckhoff's classification, regulative norms from Mayntz classification, and both authority and parts of incentive instruments (especially subcategories of sanctions and the use of force) from the classification of Ingram and Schneier. As is visible in many of the classifications, this category includes only rules imposing duties or obligations. Only prescribed rules that guide behaviour directly are included as noted by Mayntz[447]. A characteristic form, and also its caricature, is the command and control model: legal rules backed by civil, criminal and administrative sanctions monitored and enforced by a government-empowered body.

The notion of the directness of the guidance makes visible the way the category of prescriptive rules influence behaviour: by telling what action is expected from the regulated and by stipulating that negative consequences follow if the standard is not complied with. Basically an authoritative order is made. Obedience is assumed, but backed by the threat of negative consequences (not just physical or monetary sanctions under criminal law, but also any type of negative consequence attached)[448]. It has to be noted that authority, together with negative consequences, is the defining feature of directly

---

[446] This is also noted by Vedung in Policy Instruments, p. 31.

[447] Also precedents, to the degree they can be considered to communicate an authoritative example to individuals which they are expected to follow and obedience of which is backed by the threat of sanctions, belong to this group.

[448] Even though Ingram and Schneier categorize sanctions and the use of force into the incentive instrument class, they are not separate policy instruments as such. The sanctions are just one type of the negative consequences attached to actions by regulation that provide reasons for action. Even though this notion does not do justice to the analyses of punishment in social theory, it manages to explicate the larger influencing role of punishment. It must be stressed, however, that punishment is not merely a technical task of specialised institutions as David Garland in *Punishment and Modern Society* brilliantly illuminates while discussing the major social theories of punishment.

influencing prescriptive rules when analysing the external influencing mechanism[449].

It is important here to keep separate the two ways in which regulation can provide guidance for behaviour as an external constraint; direct and indirect[450]. The legal norms are used as an example. When the legal norms guide behaviour directly, the objects of regulation are told (in the general standard for behaviour communicated) how to behave or what action to abstain from and that certain typically undesirable consequence might follow in the case of non-compliance. In terms of Raz: "it guides action determinately by expressing the intention that it shall be performed and stipulating a generally undesirable consequence to follow when it is not performed" or vice versa (with the consequences remaining undesirable)[451]. Certain options that the object of regulation consider in making the decision to alter her behaviour are prohibited, commanded, or made subject to permission. This is how duty imposing norms influence.[452]

---

[449] In the widening of the sanctions (fines, imprisonment) to all negative consequences lies the difference to those categorisations, such as that made by Vedung, Policy Instruments, p. 31, where authority is not necessarily accompanied with sanctions at all. Sanctions become essential when keeping external constraining analytically separate from the intrinsic predispositions shaping mechanism, where authoritative arguments as such (without consequences attached) can influence behaviour.

[450] This separation has been made especially in relation to law. See, e.g., Raz, On the Functions of Law, p. 283-287, who discusses the issue in terms of determinate or indeterminate guidance. In the Finnish discussion this separation has been raised, e.g., by Wikström in Ohjaaminen oikeusnormeilla, p. 399-407 and in *Oikeus ja talous*, p. 5-6. Bear in mind that this is a pure normative analysis where the effort is to find the reasons for action provided by regulation as such. There are other, possibly even more important motivational factors than the reasons for action (certain consequences) provided by regulation, but they are excluded for analytical purposes.

[451] Raz, On the Functions of Law, p. 285.

[452] Note that the communicated standard is general in two ways as H.L.A Hart explains in *The Concept of Law*, p. 21: it indicates a general type of conduct and applies to a general class of persons who are expected to see that it applies to them and to comply with it.

In the light of the decentring thesis adopted in this study it is obvious that the direct form of influencing is not something only a state can use. Command and control is not only a typical form of governmental prescriptive rules. Others also essentially use authoritative norms to communicate general standards of conduct and try to tell directly how to behave. The discussion of self-regulation is illuminating. As pointed out in regulatory theory, self-regulation is often just self-administrated command and control: in the simplest form it involves an organisation or association (e.g., lawyers association) developing a system of rules (e.g., lawyers ethics) that it monitors and enforces against its own members or, in some cases, a larger community[453]. Non-legal prescribed rules include instruments such as guidance, industry codes of conduct and best practices which are not legally binding.

However, not all norms are coercive orders. Even the legal rules come in variety of forms[454]. In addition to guiding behaviour directly, they can also be used, for example, to create or alter private rights[455]. These power-conferring (constitutive) rules define and thereby constitute activities that could not otherwise even exist. They do not tell directly how to behave and attach certain negative consequences to non-compliance. However, they still regulate (guide) behaviour

---

[453] Baldwin and Cave, *Understanding Regulation*, 1999, p. 39; Black, Decentring Regulation, p. 123.

[454] There is a variety of forms prescribed norms may take. Even though the classic stereotype is the detailed legal provision, written norms can vary across different dimensions: scope (e.g., whether they cover certain functions or market sectors), substance (e.g., whether they are specification, process, target, or performance standards), status (legal or some other), whether sanctions are attached, structure (precise or general, simple or complex, clear or opaque), and place or origin (state and non-state actors on different levels in coordination or in no formal integration). Black, Enrolling Actors in Regulatory Systems, p. 67-68.

[455] Market actors can also use private law powers to create more limited private rights, for example to demand collaterals, or require certain disclosures, assuming they have the market power to do so (and the law will support their efforts). Black, Mapping the Contours of Contemporary Financial Services Regulation, p. 4.

as has been noted in the philosophy of the law[456]. The mechanism by which they guide behaviour is different, though.

Their influence mechanism is indirect[457]. Instead of telling persons how to behave or not to behave and possibly threatening some sort of punishment or some other negative consequence if they deviate from the standard, the law can stipulate certain legal consequences (desirable or undesirable) to follow upon the performance of or abstinence from certain action with the intention that these legal consequences will affect people's decisions to perform the action that is permitted as such. When this type of behaviour directing is used the regulator does not interfere in the behaviour of a person or some other actor directly; the behaviour is neither prohibited nor required. There is wider choice left to the regulated agent[458]. The regulated

---

[456] This definition of constitutive rules is taken from Schauer, *Playing by the Rules*, p. 6-7, who acknowledges the usefulness of the distinction between regulative and constitutive rules for some purposes, but points out that the distinction is also misleading. Many constitutive rules also regulate. They first define the behaviour and then regulate it. Thus, regulative rules regulate behaviour within the activities defined and thus constituted by the constitutive rules just as they regulate behaviour within environments whose definition is less rule-dependent. In terms of Schauer, idem, regulative rules not only govern antecedently *existing* behaviour but also antecedentely *defined* (and thus constituted) behaviour.

[457] Raz, On the Functions of Law, 1973, p. 283-284 uses the term indeterminate guidance. The indirect guiding function is on the basis of the reflexive or procedural law as a strategic model developed by Gunther Teubner (of the long line of work, see, e.g., After Legal Instrumentalism?, p. 307) and is emphasised in regulatory studies by Julia Black in Proceduralizing Regulation. Part I, p. 598, and in Decentring Regulation, p. 125- 126. In the Finnish discussion this has been pointed out by Wikström in *Oikeus ja talous*, p. 5-6 and in Ohjaaminen oikeusnormeilla, p. 402-403.

[458] Naturally the object of regulation also has certain freedom of choice also in relation to the direct influence mechanisms. The difference is that when the law is used in a direct sense the regulator has made the choices on behalf of the regulated. The object of regulation has only the freedom to choose whether or not she complies with the law; if not, then she is a lawbreaker and certain legal consequences might follow. The law directly interferes with the behaviour. When the law is used in an indirect sense the desired behaviour is not directly commanded or prohibited. Instead, the behaviour is incited and, despite of not complying with the regulation, one still can be law-abiding. The difference is in

action is allowed, but made conditional in the achievement or avoidance of certain consequences. The law just makes the guided action more or less desirable than it would otherwise be, i.e., without the consequences attached to action in the regulation. Behaviour is incited rather than commanded.

> This is the distinction of prescriptive rules, i.e., rules that are used to guide, control or change behaviour of agents with decision-making capacities (as separated from descriptive rules that only describe or explain the world), into instructions and mandatory rules made by Frederick Schauer[459]. The prescriptive rules to which Schauer refer to as instructive rules or rules of thumb, provide only useful guides for the routine case and are optional in the sense that they apply only if an agent wishes to succeed in the pertinent task and considers that it is likely that by following the rule of thumb the desired result will be produced. As Schauer points out, mandatory rules that exert normative pressure by their mere existence are not for the agent wholly optional[460].

What this means is that only those prescriptive rules that influence behaviour directly belong to the category called prescriptive rules.

---

the degree of constraint set for the objects of regulation or power the regulator has invested in the attempt to regulate; the extent to which a tool restricts individual behaviour as opposed to merely encouraging or discouraging it. The separation follows the amount of legitimate or permitted alternatives (choices in a decision-making situation), not the extent to which behaviour is prevented in reality.

[459] Schauer, *Playing by the Rules*, p. 3-6

[460] Schauer, *Playing by the Rules*, p. 4. Despite of this wider usage of the term prescriptive rules, where also instructive rules or rules of thumb are covered, I still prefer the term prescriptive rules over mandatory rules mainly because the term prescriptive connotates also the presence of somebody prescribing as emphasised by Raz in Reasons for Action, Decisions and Norms, p. 481. Thus, it involves the intentionality element central to the conception of regulation in this study. Note that while pointing out that 'prescriptive' is often used instead of 'mandatory', Raz himself prefers the term mandatory over prescriptive for the very same reason (idem.). It is these prescriptive/mandatory rules of which moral, legal and political philosophers have been most concerned as Raz points out in Reasons for Action, Decisions and Norms, p. 481.

Prescribed rules that confer powers, set up or alter processes, define roles or in any other way guide behaviour indirectly do not belong to this category. They do not prohibit or command certain behaviour directly. The regulated action is allowed, but made conditional in the achievement or avoidance of certain consequences. They belong to the other relevant instrument categories.

At the same time this means that only prescriptive rules can influence behaviour directly even though there are classes of prescriptive rules that influence behaviour indirectly. All the other types of instruments can influence behaviour only indirectly. Prescriptive rules that provide indirect guidance are not the only possible instruments that influence behaviour indirectly. There are also classes of instruments that provide indirect guidance beside prescriptive rules that do so.

If were not the direct forms of influencing purely at the use of the state, neither are the indirect guidance methods confined to the state as pointed out by Julia Black with her decentring thesis[461]. All those attempting to regulate behaviour, whether state-centric or not, can induce behaviour instead of ordering. A common theme is that the regulator does not interfere in the behaviour of the regulated directly; the regulators create the conditions in which firms, markets and communities steer themselves, but into the direction that the regulator wants them to go[462]. This is visible especially in the other types of instruments that provide indirect guidance beside prescriptive rules that do so.

With these other indirectly guiding instruments the reasons for action provided by the regulation are especially mediated by the context of the decision-making. The consequences attached to actions are only partial and incomplete reasons for action (incentives) among

---

[461] Black, Decentring Regulation, p. 111-112.

[462] The regulator can influence the availability of choices (direct the decision-making process), e.g., by changing the mental desirability of certain choices, or by changing the physical or virtual environment to constrain the choices. Both of these methods influence as external constraints in the sense noted above. In addition, the regulator can influence behaviour by designing decision-making procedures and institutional structures.

a wide variety of complementary factors in various circumstances even from the objective perspective[463]. The mediators of the guiding influence of the regulation are the constraints on individual behaviour beside the formalised general standards of conduct like legal norms made explicit, e.g., by Lawrence Lessig: market via prices, social norms in the form of informal social control of the community or the inner moral norms of the self, or technology that defines the set of possible behaviours both in the physical and the virtual spaces[464].

The *economic or market-harnessing* instruments include, similar to the Black's classification, incentive (excluding sanctions and the use of force) and parts of capacity tools (those providing economic resources) from the classification of Ingram and Schneier. Naturally the economic instruments (carrots) from the classification of both Eckhoff and Vedung, and the treasures using category of the resource based classification are also included, with the exception of the in-kind (physical) instruments being excluded and forming a separate category. The economic instruments influence the considerations of what kind of behaviour is economically rational. Their intention is to make it cheaper or more expensive in terms of money, time, effort, and other

---

[463] As Raz, On the Functions of Law, p. 284, notes whether the consequences are reasons for performing or refraining from the action depends on these changing factors. In the case of direct influencing, the stipulated legal consequences are independent reasons for action in themselves.

[464] Lessig, *Code and Other Laws of Cyberspace*, p. 86-90 and appendix. The purpose of my study is not to analyse the relationship between regulation and social norms, market rules or even technologies. This mainly because they are extensive research areas in their own and I do not have much to contribute in those discussions. The interest is in the tools that can be used to harness them for regulatory purposes. By noting that these forces can be harnessed in the intentional process of attempting to alter the behaviour of others and looking for the tools to do so, only the surface of the discussion concerning the interaction of the different constraints on behaviour is scratched. A simple understanding without deeper knowledge is enough to see the tools regulation might use to harness them.

valuables to pursue certain actions[465]. The instruments either hand out or take away resources conditionally (provide economic incentives for behaviour), or harness markets for regulatory purposes (by channelling competitive forces to particular ends)[466].

Included are tools like taxing, deployance of grants or subsidies, tradable permits, charges and regulation by contract. An example can be found from tax relieves for those acquiring broadband access to the Internet. The relief is a legal consequence attached to the action of acquiring the connection that provides reasons for taking that action; it lowers the costs of acquiring the broadband connection. The guiding (the legal consequence attached) is mediated by the economic (price) incentive. The market mechanisms are harnessed for regulatory purposes.

Even though the economic instruments might seem more prohibitive (preventive) than regulation, for example in cases where a sanction of breaking the regulation is relatively low but the excise tax levied on the same line of action is enormous, in principle, regulation is more constraining than economic instruments (however expensive they make the behaviour) because one is not allowed to take the action if it is prohibited (an offence or a breach of duty is involved) but, in the case of economic instruments, the action is

---

[465] Economic instruments affect by raising or lowering the costs of engaging in a particular activity and thus provide powerful economic incentives to undertake the desired behaviour or to avoid the undesirable behaviour as pointed out by the OECD Working Group on Regulatory Management and Reform in *Regulatory Policies in OECD Countries*, p. 138.

[466] The basic rules of markets outlined in microeconomic theory have an important influence on human behaviour; markets create incentives for people to behave in particular ways, especially via prices. This is what economics, with its mathematically precise theories (price and game theory), tell us. The market governs individual behaviour by setting the substance and constraints upon individual choice. This theoretical construction makes it possible to alter the included rules intentionally; economic theory makes it possible to harness market rules in the process of regulation. Naturally the markets are not 'unregulated' because they are essentially based on the private law rules, application of the law in courts and the enforcement processes (e.g., of private contracts).

allowed although it will cost a considerable amount of money to perform it[467]. The separation follows the amount of legitimate or permitted alternatives, not the extent to which behaviour is prevented in reality. The question is about the extent to which a tool restricts individual behaviour as opposed to merely encouraging or discouraging it as noted above.

Many of the *informative or pedagogic* (social norms harnessing) instruments simply seek to inform and enhance consumer choice, but some of them are more explicit in seeking to change behaviour. They are based on attempts at "moral suasion" and are generally found where the behaviours sought to be modified have substantial externality effects[468]. The informative and pedagogic instruments harness social norms for regulatory purposes by making symbolic or hortatory proclamations to influence perceptions of values or learning to reduce uncertainty. It combines the sermons, the pedagogic and informative instruments, instruments utilising the nodality of regulators and the social norms and sanctions from the above classifications. The informative instruments contain information, transfer of knowledge, the communication of reasoned argument, persuasion and education and are often performed in the forms of information or education campaigns.[469]

Even though Black in her general taxonomy includes also legal rules into the tools that can be use to alter social norms, in addition to education and information, the use of peer pressure to ensure that others act in accordance with certain standards of behaviour, and

---

[467] This differentiation is made by Vedung in Policy Instruments, p. 32-33. This is also the approach in traditional legal theory. See, e.g., the differences between a tax and a punishment for a crime such as a fine discussed by Hart in *The Concept of Law*, p. 39.

[468] OECD, *Regulatory Policies in OECD Countries*, p. 139

[469] Note that the sanctions attached to social norms (from the outside, i.e. others in that group) are not set by the regulator. However, the regulator can try to alter the sanctions but it is just part of the issue of changing social norms.

simple blacklisting and their publication[470], they do not do so in the *external constraining* sense. They can be used to alter social norms, but only as information *on* the rules, i.e., as a metainstrument by the means of expression and justification. This was analytically separated above from information *as* a policy instrument. The justifications given for a written rule, or an economic, technology, process harnessing, or even the informative instrument, and the way the rules are expressed also involve persuasive element, but this should be separated from the information as a separate externally constraining policy instrument.

It needs to be acknowledged that the informative or pedagogic regulatory instruments (information *as* a policy instrument) can also influence behaviour either as external constraints or by shaping intrinsic predispositions. As pointed out by Etzioni, they both inform about the costs, benefits and constraints involved in certain action (e.g., of the harms of smoking) and empower the relevant social group to direct pressure towards the deviator (e.g., shaming, shunning), and thus affect the consideration of the content and intensity of numerous particular predispositions in a decision-making situation, and help people to form (and re-form) the self by profoundly influencing their identities, their worldviews, their views of themselves[471]. A good example is advertising. Advertising is not strictly informational and thus not only affects behaviour by affecting considerations of costs and benefits (and, as a rule, do not shape preferences; i.e., advertising is considered to affect as an external mechanism) but also a means of subconsciously affecting people's preferences and thus a form of persuasion in the terms of Etzioni[472].

If were the intrinsic predispositions shaping as a way for the regulation to influence behaviour mediated with the non-rational process of internalisation and persuasion by which people adhere to them, similar non-rationality can also be seen in relation to the external

---

[470] Black, Mapping the Contours of Contemporary Financial Services Regulation, p. 7.

[471] Etzioni, Social Norms, p. 163.

[472] Etzioni, Social Norms, p. 169-170.

influence mechanism of informative and pedagogic instruments (information as a policy instrument, social norms harnessing instruments). As pointed out by Weiss in her analysis of public information as a governance tool, while acknowledging that people react to the information provided by the regulation, it is not necessary to adhere to the rational choice model of decision-making[473]. Although there is some truth, and strong explanatory force, in the simple view that people make decisions by adjusting automatically and continuously to information that bears on any consequences of their actions, much of the time full information is not available to all or feasible to gather, and all individuals are not perfectly responsive even to the information that is available. In addition, the information may affect both the explicit and implicit cognition, i.e., people may still respond to information in the world around them and connect it to their subsequent behaviour even though they might not be making deliberate, conscious calculations about what they do.[474]

Social norms and markets are, to certain degree, clear objects to be harnessed in the process of regulation. Their relation to regulation and especially the law is not even questionable. There are established research areas considering both of them: those leaning on sociology (legal sociology, law and society, etc.), and those leaning on the economics (law and economics with its different schools but a common paradigm). There are even considerations of some sort of a common ground between the study social norms in the disciplines of law-and-society and law-and-economics[475].

---

[473] Weiss, Public Information, p. 218-219 and 227-228.

[474] This is held also by Tala, *Lakien vaikutukset*, p. 18.

[475] This is what Etzioni labels as 'law and socioeconomics' in a longer line of study. See, e.g., Etzioni, Social Norms, 2000, p. 158. Etzioni argues in Social Norms, p. 157-179, that the common grounds are found by combining the themes of (a) whether social norms affect individual behaviour merely as external constraints or whether they also shape people's intrinsic predispositions, (b) whether preferences are considered predetermined or assumed to be modifiable as a result of internalisation and persuasion, and (c) the ways social norms are themselves formed (merely via rational choice or

This is not the case with technology. It is not a part of mainstream regulatory literature and not typically considered even in the socio-legal studies[476]. Thus, the situation is not even as clear as with the role of social and market forces in the process of regulation, which is widely acknowledge even though not widely known or deeply analysed. This is why the role of technology has to be taken a little bit further, even though the interest in here is not on how technology as such operates similarly to treatment of market rules or social norms (the purpose is to analyse intentional attempts to alter them).

Technology can be understood at least in three different ways[477]. In the *first*, used by most legal professionals as well as regulationists, technology is considered as something requiring knowledge of technical details; it is just about machines and devices that have a really

---

also through historical transmissions). It has to be noted that law and economics scholars dominate the contemporary study of social norms in the legal academia. The emerging discussion of the combined study of social norms between the law-and-society and law-and-economics scholarship, and the overall increasing interaction between these disciplines cannot be taken further in this study. For overviews, see Law & Society Review, Vol. 38, No. 2, 2004

[476] Economics (together with law and economics) based theories of regulation do not see the role of technology. It is hard to combine the regulatory force of technologies when preferences are considered as stable and exogenous. But social (or socio-legal) theories of regulation are no better. Even though they consider preferences as endogenous, they miss the role of technology in concentrating only on human behaviour. This means that when considering the role of technology in the process of regulation, the focus needs to expand from the interests and actions of humans (the main issue in regulation) to the constraints and possibilities of the relevant technologies as such.

[477] The definitions in their basic form are taken from MacKenzie and Wajcman, Introductory Essay: The Social Shaping of Technology, p. 3-4. A generalized speaking of 'technology' as a coherent and material thing with homogeneous and undifferentiated characteristics, and not just about certain technologies (techniques) is aligned with technological determinism (due to reification, i.e., treating an abstraction as a material thing) that I am trying to avoid and somewhat against to due to its role in mystifying and detaching technology from the interaction of social forces. But it is very difficult to avoid reification if one wants to use at least some kinds of generalizations. See Chandler, *Technological or Media Determinism*, part "Reification".

vague relation to the society. Technology is just a tool or a device. This narrow definition of technology separates it from the sphere of interest of legal or regulatory research and leaves the studying of it to the technical professionals. It is something marginal outside the disciplines of regulation and law[478]. In the *second*, people are brought to the scope of the conception technology. Even a computer is not meaningful without a user and the purpose of use. This increases the interestingness of technology to legal and regulatory research because people's behaviour in various situations is the main target for (legal) regulation.

In addition to machines, devices, their users and the needs of these users, there also is a *third* level in understanding technology. It includes people's knowledge and doings with relation to these machines and devices. Technology is knowledge; the know-how to use them, repair them, design them and make them. It is also about the holder of this knowledge and the right to use it. This adds the organizations and the institutions using this information as well as their architecture and interdependence to the definition of technology. Technology, in this sense, includes also the power over the economic and productional processes as well as the consistent social circumstances of the technological systems. In this sense, technology is knowledge and transmission of information (communication); it is one of many generalised systems of communication together with politics, economy, science, art, law, religion etc. also known as social functions[479].

---

[478] This is how expertise in copyright and patent law, the few areas of law that directly deal with technology and from where legal professionals get their view of the emergence of technology, has traditionally been seen by legal professionals – marginal and requiring knowledge of technical details. The informatisation of life, however, brings these disciplines into the forefront of legal expertise and may require substantial changes to the system of intellectual property rights as Annamari Turunen argues on the basis of an comprehensive analysis of the legal architecture for innovations.

[479] The development of society is understood to consist of the evolution of and interaction between these generalised systems of communication. See, e.g., Sand, A Future or a Demise for the Theory of the Sociology of Law, p. 56-60.

This third understanding is the way 'technology' is to be understood in the regulatory arena; not just as machines and devices, not even when combined with the users and different uses. It includes also people's knowledge and doings with relation to the machines and devices and the holder of this knowledge and right to use it. Technology is a product of understandings of the physical and human environment together with the ability to employ, manipulate or alter it[480].

> Technology, in this sense, is neither deterministic (technology as sufficient or, at least, necessary condition determining widespread societal or behavioural changes) nor voluntaristic (individuals choose the tools they use and have perfect control over them; the direction of technological development is chosen by individuals). Technology is a system where both technological and social actors operate[481]. Technology is not deterministic because it requires human

---

[480] This is how Black in Mapping the Contours of Contemporary Financial Services Regulation, p. 6-7, refers to technologies: as specialist understandings of and ability to employ, manipulate, calculate, measure or alter the physical, economic or human environment and the products of that understanding. The place of 'technologies' in regulation is thus that they are the understandings of the world on which the design of both the problem that regulation seeks to address and the solution to that problem are built, and it is suggested that they form an important part of the strategies that are used in attempts to modify behaviour. At their simplest they may be the design of the physical environment or its virtual equivalent.

[481] There are different approaches to the ways these actors operate. For an overview to the wide discussion about the role of technology see, e.g., Hosein, *Regulating the Technological Actor*, p. 55-68, with extensive references. *Anti-essentialists* believe that researchers must listen to the social interpretations of the technological; the technological actor does not speak. *Social constructivists* argue that we must open the discourse and look at the granular details of the construction of the technological to see that it is a social and technological construction. Those who support the *social shaping* approach say that social actors may affect the construction of the technological; although the technological actor may also refuse to work. *Technological momentum* notices that systems may radically change due to social and technological shifts in the environment. Common argument in these approaches is that the technological actor, in fact all actors, *may* resist interpretation, construction, and shaping; the technological actor can also object to being spoken for, and can object to being translated (Hosein, *Regulating the Technological Actor*, p. 66)

> interpretations and action. Yet, neither is voluntarism the answer as technology objects to human action[482]. Technology has a role to play, but also represents the interests of its developers and is therefore socially shaped.

*Technology-harnessing* instruments utilise the ability to shape the physical or virtual environment. They facilitate or hamper certain operations in the actor's environment. Examples are buildings, devices, and other physical constructions and virtual counterparts.

Technologies define the set of possible behaviours both in the physical and the virtual spaces; they are part of the built environment of the problem regulation seeks to address together with the solution to it[483]. It is within the boundaries set by the technologies where social norms and markets can be harnessed and where regulation operates. Thus, technology sets the possible behaviours and regulation determines which behaviours predominate.

---

[482] According to Bruno Latour technologies may disrupt human action through objecting; technologies can force us to renegotiate our paths and goals and so restructure our behaviour. Latour, When Things Strike Back, p. 115-117.

[483] I refrain from using the 'Lessigian' term of architecture (the built environment; the way the world is; the combined constraints of physics, nature, and technology that in the aggregate define the contours of the places where human behaviour occurs and the things through which it is expressed) in this context. The message that is included in the term is by no means clear or obvious; it is difficult to see what is referred to when speaking about 'architecture' since it has so many different usages. I do acknowledge that it necessarily not always is technology, in any of its definitions, that constraints behaviour when speaking about constraint of the built environment. There are also non-man-made constraints, such as the examples used by Lessig in *Code and Other Laws of Cyberspace*, p. 236: gravity and the limit to the speed human beings can travel. The nature and physics set the limits of our behaviour. However, these constraints of nature and physics are present in the technologies and technological systems; they are developed on the basis of understanding or and ability to manipulate this physical and human environment together with its virtual version. Technology is a product of this understanding and ability to manipulate. In addition, these non-man-made constraints are not all that important in the context of information security; we build the physical and logical constraints of behaviour in relation to information and information systems.

Different kind of rules can be embedded into technological systems. In the virtual environment the decisions of programmers about software design set the rules for behaviour just as the decisions in the physical environment, e.g., of road planners and bridge builders, control what can be done and where. Harnessing these rules and their designers in the intentional process of regulation has become possible with the developed understanding of this process. A clear example of the kinds of rules that can be embedded into technological systems is provided by the use of filters to block access to web sites containing child pornography[484].

Technology, as a constraint of behaviour, is as important as social and economic forces in that it affects the activity of regulating (sets the basic limits regulation can achieve) and can be deployed in that activity (alteration of the technology for the purposes of the regulator). But the technology harnessing instruments are different. Where there are actors or forces setting up and enforcing (executing) the constraints that harness social norms and markets, the constraint of technology, even though often generated (build) by people, is self-executing[485] and is imposed without any requirement of subjectivity[486]. Thus, the technology harnessing instruments are not only automatic,

---

[484] Such systems are widely used by the ISPs in Europe on a voluntary basis and often in a self-regulatory manner in cooperation with the police as the recent report to the Ministry of Transport and Communications explicates. The report titled as "Selvitys lainsäädännöllisistä esteistä ja muutostarpeista rikollisen ja lapsille haitallisen sisällön estokeinojen velvoittavuuteen liittyen", LVM Hanke 42976, Dno. 1394/92/2005, is available in Finnish at http://www.mintc.fi/oliver/upl973-Railaksen%20selvitys.pdf [19.12.2005]. It is part of the consideration to implement similar voluntary self-regulatory filtering system also in Finland.

[485] There is no individual or a group enforcing the constraint; the enforcement is embedded in the design. For example, bars over windows, once installed, prevent entry without the continuing intervention of other individuals. For the general argument see, e.g., Reidenberg. Lex Informatica, p. 565 and Berman, Cyberspace and the State-Action Debate, p. 5.

[486] The constraints are enforced without the need to be internalised by the individual experiencing the constraint. This is the argument of Lessig in *Code and Other Laws of Cyberspace*, Appendix, p. 236-239, especially endnote 8.

in the sense that they do not have to be separately enforced, but also more effective, because technological constraints work whether or not the subject knows they are working (for other instruments to work, the objects of regulation need to know at least something about them)[487].

Regulators, whether state centric or not, need to understand the role of technology because the possibilities provided by it help to drive the rulemaking process. It is often a more efficient way to enforce rules than sanctions (due to lack of subjectivity requirement; it enforces without the need to be internalised) and even more importantly, it is an invisible way of doing this; there is no political risk involved because the regulation is being done via alteration of physical environment. The regulatory function of technology is normally not made explicit; it appears to be just the way things are[488].

Beyond this, the technology harnessing regulatory approach does not offer an easy tool for solving current and future problems. The providers and administrators of the relevant technologies are diversified; often there is no single party or a coherent group. Finally, the regulatory effects of technology harnessing instruments are often unintended, not least because the ways it affects is not general knowledge. For technology typically is developed by technicians who tend to be concerned by not much more than technical efficacy. Its regulatory function has not been of much interest. In addition, the regulatory mechanisms implemented in technology can be circumvented, just as physical barriers can be broken through[489].

---

[487] This is also noted by Vedung in Policy Instruments, p. 43, according to whom "in-kind economic instruments are more constraining to recipients than in-cash instruments".

[488] This is noted, e.g., by Boyle in *Foucault in Cyberspace: Surveillance, Sovereignty and Hard-Wired Censors*, Heading "Conclusions" second paragraph, and by Lessig in *Code and Other Laws of Cyberspace*, Appendix, p. 236.

[489] For example, when the efficacy of technological protection of copyrighted works was demolished with the new circumvention means, other regulatory instruments needed to be taken into use to protect the technology as the famous cases of anti-circumvention regulation in the DMCA (Digital Millenium Copyright Act, the EUCD (European Union Copyright Directive)

Combining regulation (especially law) and technology is not a straightforward task. Normative expectations and legitimate decision-making by law together with democracy and the rule of law are still vital parts of the coordination of modern societies. However, together with the complexity and flexibility of the 'new' technologies and their rapid change, also the steering of technology as part of the coordination of a society is at change. Possibilities for external intervention seem to diminish. At a time of rapid and intensive change it is difficult to give legal and other normative expectations that usually are based on earlier experiences and require at least some degree of stability. The complexity of systems and the speed of development hinder the construction of adequate legal versions of reality (legal visions of certain issues)[490]. Legal structures easily become inadequate and insufficient.[491]

The passing of normative expectations and legitimate decision-making is further hindered when the technology is seen as machines, devices, and software that require understanding of technical details, thus alienating regulators from technology, especially those not familiar with it. However, seeing both regulation and technology as communication helps to connect their development. They are both one of many generalised systems of communication together with

---

and the Conditional Access directive illustrate. For an excellent overview of the regulatory background, see the analysis of Peter Drahos and John Braithwaite in *Information Feudalism*, p. 184-186, drawing on data from interviews of key informants of the rise of the TRIPS agreement (Trade-Related Aspects of Intellectual Propery Rights).

[490] The complexity of operational environment in information technology is not restricted on mere entirety of software and hardware. The people, the institutions and the organisations associated with it make the operational environment even more complex. This extensive definition of technology – technology as a system where software and hardware are united with people, institutions and organisations to work as a seamless network – makes the information infrastructure so comprehensive that it is really difficult to piece it together.

[491] More about the effects of the acceleration of technological development to the law as communication see Sand, A Future or a Demise for the Theory of the Sociology of Law, p. 57-59.

politics, economy, science, art, and religion etc. also known as social functions. Technology in the widest sense (as communication) helps to master the complexity and to adapt to the flexibility, by showing its relationships with and effect on society, law, economics, politics, science, etc. This comprehensive definition of technology makes more easily visible e.g. the possibilities of regulation to adapt to the changes in society by reforming the regulation of institutions to favour certain solutions.

Technology affects society, even profoundly. So does law and other regulations. And society affects technology, law and regulation in general. This is clear even for a legal professional. The relationship between technology and regulation just seems to become distorted. It is misleading to consider that relationship similarly direct as the relationship between technology and human behaviour. There are social practices and institutions mediating them[492]. This makes the indirect regulation via technology difficult for the government because technology is less and less in direct ownership or control of government or some other party. With privatisation (increasing contracting out government services and use of government sponsored enterprises instead of government corporations or direct government), general use of intermediaries in regulation and decentred regulation in general, the indirection multiplies. Regulating behaviour via changing technological habitat has been done, to increasing degree, by using several mediating forces and actors. In relation to 'new' ICT this is especially problematic due to the still existing basic starting point of laissez-faire and heavy competition.

*Process-based* instruments are a bit of a special case. Procedures, participation, and institutional design are the common solutions advocated currently; the dominant call is to develop procedures and

---

[492] The main mediator between the law and the technology is the market, but also governments and participants of civil society (e.g., families, civic organisations, workplaces, religions, communities, and etc.) are between law and technology as Joseph H. Sommer points out in his critical analysis of the technologically systematised bodies of law such as "cyberlaw", *Against Cyberlaw*, heading "The Perils of Cyberlaw".

institutional structures that will enhance deliberation and enable participation[493]. These instruments are so named because they structure or require the development of organisational or decision-making processes in a way that attempts to ensure that appropriate outcomes are reached even though the appropriate outcomes or the decisions to be reached are not necessarily specified.[494]

The procedural regulations, even though they do not typically determine the results the participants in the procedures ought to achieve (they are formed during the action of the participants of the procedure), guide behaviour indirectly because the procedures are being set in place, and for example, participation enabled or deliberation enhanced, with the objective of achieving certain types of results (even though they might not specified)[495]. For example,

---

[493] Black, Proceduralizing Regulation, p. 597.

[494] Black, Mapping the Contours of Contemporary Financial Services Regulation, p. 7. Note the similarity to above mentioned classification of Mayntz in The Conditions of Effective Public Policy, p. 139, where procedural rules are set for the cooperation and/or resolution of conflicts between two or more parties without trying to control the outcome of these processes directly. In addition, see also Tala, *Lakien vaikutukset*, p. 176, who includes the regulation of legal processes into his classification of legal regulations. The process-based instruments are also similar to the procedural programs (prozedurales programm) in the German classification of different political programs made in implementation research. Reference is made on Tala, *Lakien vaikutukset*, p. 159-165, because the originals are in German. The original German text to which Tala refers to are Mayntz R (1983) Zur Einleitung: Probleme der Theoriebildung in der Implementationsforschung, in Mayntz (ed.) *Implementation Politischer Programme II*, Westdeutscher Verlag, pp. 11-13 (7-24), and König K and Dose N (1989) Klassifizierungsansätze staatlicher Handlungsformen, Speyerer Forschungsberichte 83. Procedural regulation is also defined as a form of state centred regulation by Jarass in Regulation as an Instrument of Economic Policy, p. 79.

[495] As pointed out by Black in Mapping the Contours of Contemporary Financial Services Regulation, p. 7, the institutional structure or procedure might also be the goal in itself, in addition to being the means for achieving a certain desired goal or policy objective. In addition to using processes to achieve certain material decisions or the quality of products, they can be used self-sufficiently (the goal being or that the process itself respects certain values). Processes can also be used more explicitly; demanding that they are put in place, but at the same time specifying the type of outcome that should be

when the law is used to affect behaviour structurally, by constituting new institutional structures and by strengthening existing institutional structures, the content of the behaviour typically is of less interest and the desirability of certain choices is not increased or decreased as such[496].

Their influence is indirect in another sense. They guide behaviour by demanding the objects of regulation to deliberate, to use a systematic approach in doing things, to take notice of the interests of certain parties (by enabling them to participate or by requiring them to be heard), or to make the processes public and open. Whereas other instruments add consequences to be considered, the process-based instruments alter the context of the decision-making[497]. They concern organisations and procedures, the redistribution of power and competences; they create or alter procedures for the support of the decision-making and rely on the decision-making and operational systems together with their design.

Examples are processes that ensure a systematic approach to controlling and minimising productions risks (like development

---

reached and imposing liability if this does not happen. As Jarass notes in Regulation as an Instrument of Economic Policy, p. 79-80, in most cases procedural regulations are combined with substantive regulations which prescribe desired conduct. Process-based regulation may also take the form trying to balance opposing forces or to solve conflicts through the design of processes.

[496] An example, of the law strengthening existing institutional structures, but not constituting them is the relationship between law and markets; the market economy needs at least private law to operate but it can exists in spite of the legal system. In addition, one of the functions of law is to strengthen existing social norms but social norms are seldom constituted by the law. An example of the role of law in constituting institutional structures is taxing. Wikström, Ohjaaminen oikeusnormeilla, p. 400.

[497] Whereas the indirect guidance provided by other instruments is mediated by the other constraints on individual behaviour (than formalised, duty imposing rules), the influence mechanism of the process-based instruments is mediated by the structures of the institutions that implement or enforce the constraints (also the constraint of formalised rules), or their decision-making procedures.

methods)[498]. Also the institutional or organisational forms used belong to the process-based category[499]. An example could be a regulator (e.g., state, industrial pressure group, labour market organisation, consumer group) trying to enhance the public interest by requiring corporate boards of directors to include a certain number of representatives that are not dependent on the firm, in the hope that the boards would then by themselves decide to act more consistently with the public interest. This strategy could be used instead of regulation directly requiring corporations to do certain things that would benefit the public interest.

In order to make processes a distinctive category, they have to be separated from the tools that can be used inside these processes. For example, a party in a cooperative process (e.g., required negotiations) may try to persuade the other party by threatening to use more coercive means (e.g., direct action). Information is typically used inside processes (especially in participatory) to influence others. The regulation through the design of processes does not cover these (they are distinctive tools in themselves). Instead, it concentrates on the surrounding organisational and decision-making environment and their structures.

---

[498] As the OECD Working Group on Regulatory Management and Reform notes in its review of regulatory reform in OECD countries, procedural rules can also be used in a manner where the regulator requires businesses to develop processes that ensure a systematic approach to controlling and minimising production risk. OECD, *Regulatory Policies in OECD Countries*, p. 136. The term process based instrument is also used in this report for this type of regulation.

[499] In the first comprehensive Finnish analysis of information security norms Saarenpää and Pöysti, *Data Security and Law*, heading 13.4, classified them into nine groups of which one is the norms regulating the functions and structure of organisations whose task it is to promote and provide data security.

## 3.4 Packaging influence mechanisms - the interaction of instruments

Regulatory instruments are not used in isolation. They come in packages and they interact in many ways[500]. A basic notion is that regulatory instruments are chronologically sequenced, i.e., they are selected in a certain time order. In addition, the above mentioned classifications of instruments on the basis of coerciveness has an implicit hypothesis according to which, other things being equal, the more coercive the tool, the more effective it is likely to be[501]. The object of regulation is expected to comply better with the more coercive tools. From the point of view of the regulated this classification coordinates with the level of restriction on the margin left where to choose; the information provided by the regulator does not limit the possible choices, where as the threat of enforcement on imprisonment does. This is the degree of constraint the regulation sets for individual behaviour. This classification also underlies the notion of evaluation studies according to which problems are reacted with escalation; starting from less coercive (like information provision and education) and moving to more restrictive forms where needed.[502]

If the clear case of different instruments being used at the same time towards the same objects of regulation is omitted[503] then vertical

---

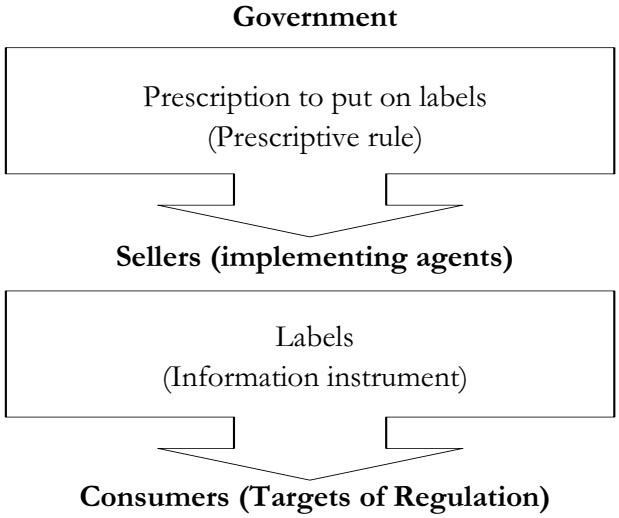
[500] Regulatory strategies rarely employ just one of the tools. The buzz word in regulation is a regulatory mix. In practice, alternative regulatory strategies are often hybrids of two or more tools; e.g., self-regulation forms a hybrid of social norms harnessing instruments and prescriptive rules (typically written rules), and contractual governance resembles a hybrid of market-harnessing instruments and prescriptive rules.

[501] Salamon, The New Governance and the Tools of Public Action, p. 26; Tala, *Lakien vaikutukset*, p. 155. However, even information tools can involve considerable coercion if information crosses the border into indoctrination. But this coercion still is in another, more subtle level. Of this, see Weiss, Public Information, p. 220-221.

[502] Bemelmans-Videc and Vedung, Conclusions: Policy Instruments, Types, Packages, Choices and Evaluation, p. 263-264; Tala, *Lakien vaikutukset*, p. 155.

[503] This is typical horizontal packaging as defined, e.g., by Bemelmans-Videc and Vedung in Conclusions: Policy Instruments, Types, Packages, Choices and

packaging is something that has to be considered. As discussed in relation to state-centred regulation, vertical packaging is the practice of directing one instrument at the implementation of another instrument[504]. The regulator uses several layers of actors to influence the final addressees, and the higher-level actors are supposed to exert influence over lower-level actors again through the use of policy instruments. The instruments used at different levels can differ and typically do. A typical example illustrated in picture 3.1 is mandatory labelling where laws are used to force producers and distributors to provide information to the consumers.




Picture 3-1 Labelling as vertical packaging

---

Evaluation, p. 262-263. Note that this is clear only from the perspective of the regulator. Horizontal packaging of different instruments with separate influence mechanisms may shape the effects created, but in the current study this is partly omitted. In an analysis of a specific implementation of certain instruments it could be possible to analyse the effects of different horizontal packages used, but it is not feasible to go through every possible kind of package since there are so many.

[504] See, e.g., Bemelmans-Videc and Vedung, Conclusions: Policy Instruments, Types, Packages, Choices and Evaluation, p. 257-263.

The use of indirect strategies of regulation, in the sense that the instruments used rely heavily on a wide assortment of third parties (e.g., corporations, universities, other levels of government) and employ other "forces" (markets, social forces and technologies) for their purposes has increased. This has been widely recognized especially in relation to law and other instruments typically used by state actors[505]. There is nothing new in the delegation of governmental powers and duties. However, governmental actions are being shared with a wider array of third parties, and even the exercise of discretion over the use of the public authority and the use of public funds are shared[506]. However, in the light of the decentring thesis, also other than state-centred actors utilise this practice[507].

It is noteworthy that both the direct and indirect behaviour influencing mechanism and every type of instrument can be used at any of the vertical stages. The object of regulation, the party considering the consequences attached to actions and whose actions are either directly limited or made more or less desirable, just is the

---

[505] Salamon in The New Governance and the Tools of Public Action, p. 2 and 8, note this in relation to the United States. Salamon, idem, p. 5-6, and 597, also notes the reliance on indirect tools in Europe as does Moran and Prosser in Privatization and Regulatory Change in Europe, p. 10. Similar points have been made also in the Nordic countries by Mathiesen in *Rätten i samhället*, p. 90, and Niemivuo in Kansallinen lainvalmistelu, p. 10-12. Lessig in *Code and Other Laws of Cyberspace*, p. 95 makes the same notion in relation to the regulation of the cyberspace.

[506] Scandinavia is a bit of a special case in this relation. Even the central study books, such as that of Salamon, The New Governance and the Tools of Public Action, p. 5-6, note that direct government is still central in Scandinavia. The legal tradition in constitutional state especially in Finland has hindered the development towards the use of more indirect regulatory mechanism and the use of non-legal regulatory alternatives to parliamentary acts as pointed out by Jyrki Tala in Lainsäädännön vaihtoehdot – tarve ja tehtävät, p. 6, but is not non-existent as the privatizations of many governmental enterprises (former governmental monopolies) already testify.

[507] Black, Decentring Regulation, p. 111-112. This is actually also noted by Lessig in *Code and Other Laws of Cyberspace*, p. 99, but he does not elaborate it and does not seem all that worried about the problems of indirection outside the state actors.

actor at the upper lever of the vertical package instead of the final target of the regulation. However, at the same time with the increased use of third party government there has been a shift towards to use of the more indirect and abstract guidance mechanisms referred to above.

Thus, prescriptive rules (especially duty-imposing legal norms, but also other legal norms) are not the only category of instruments that can be, and typically are employed to regulate actors in mediating levels and to control the use of other instruments in the lower level of packaging[508]. The will of the original regulator can be enforced by requiring implementation by third parties or other forces through means (other than prescriptive rules) like government contracts[509], through state ownership position, by changing technology or by harnessing social norms etc.

When categorising instruments and analysing their influence mechanism it is important to keep in mind to which vertical stage in the chain of implementation the instrument is directed[510]. It helps

---

[508] This has been pointed out by Bemelmans-Videc and Vedung in Conclusions: Policy Instruments, Types, Packages, Choices and Evaluation, p. 260. Lessig also notes in The Law of the Horse, p. 511-512, that "architecture", similar to all of his modalities (law, markets, and norms), can operate at a level of indirection (when using a more strict understanding of regulation: can be employed to regulate other instruments). However he concentrates just on two: the effect of law on the market, norms and architecture, and the effect of architecture on law, market and norms.

[509] The difference between contractual regulation and prescriptive rules is twofold. Firstly, there is a freedom of choice in adhering to the contract. The directions provided in the contracts are consequences attached to the action of contracting and thus only provide an incentive – they tell what to do only when the party has adhered to the contract. Secondly, when this adherence has taken place, and the contracts authoritatively tell what to do and threat with negative consequences (damages, end of contract, etc.), the standard for behaviour provided does not apply to a general class of person. It applies only to the parties of the contract.

[510] Vedung, Policy Instruments, p. 37. It should be noted that the model of vertical packaging is neither a depiction of empirical realities, nor a normative ideal. Instead, it is a heuristic tool for analysis; an ideal type as pointed out by Bemelmans-Videc and Vedung in Conclusions: Policy Instruments, Types,

to separate the different instrument from each other even though they might seem to include similar parts. For example, legal and other prescriptive rules are part of most packages. In the case of disclosure rules, prescriptive rules require mandatory disclosure (oblige suppliers of goods to provide information to consumers on price, composition, quantity, or quality), or prohibit the supply of false or misleading information. However, they belong to the prescriptive rule category only in relation to sellers (who are required to provide or prevent the disclosure of certain information). In relation to consumers, who are supposed to get the information or not to be subjected with it, and, as a consequence, perhaps act in some way or another, they are informative instruments.

For the sake of simplicity, the issue of vertical packaging is typically disregarded in the categorisation of the instruments, and the concentration is on the consumers and other final objects of regulation in the chain[511]. The main function of the instrument is considered to be the form of guidance it provides on the final targets (be they end-users, consumers, or private or governmental organisations). Unnecessary complexity in the classification can be removed by making this simplifying assumption. This is in line with the basic starting point in the normative analysis of the ways regulation guides behaviour; the effects of regulation are caused by the actions of the final object of regulation.

However, this is too simplifying for the needs of the analysis of the regulatory capability of instruments. The heuristic tool of vertical packaging cannot be disregarded when the capabilities to influence are considered. The effects of regulation on the behaviour of the final targets of regulation are influenced by the vertical stages as implementation studies have clearly shown; implementation alters the effects created. In addition, the effects on the institutions implementing and enforcing the instruments, on their procedures and practices are no less relevant than the effects on the behaviour

---

Packages, Choices and Evaluation, p. 258.

[511] This is done, for example by Vedung in Policy Instruments, p. 37.

of the final targets[512]. This is why the vertical packages are considered in the analysis of the influencing capabilities of instruments, even though the classification disregards it.

As already noted, the instruments used in influencing behaviour indirectly are essentially vertically packaged. They often are directed at third parties that then address the final objects of regulation[513]. In addition, the regulators with their different resources and use of variety of instruments with different effects in the earlier stages of the chain affect the ways the final addressees react. The upper levels of vertical packages affect capacities the final regulators have and thus become relevant for the effects regulation is likely to have on the final addressees. This is especially important in the analysis of the implementation of the instruments.

Neither are the effects on the institutions implementing and enforcing the instruments, or their procedures and practices less relevant than the effects on the behaviour of the final targets. Under consideration are then the tasks, authorities, means, responsibilities and resources given in regulation to the implementing or supervisory bodies, how regulation changes these and how it affects the operation of the institutions. In addition to the considerations on the effects of regulation on the implementing and supervising bodies, included are also considerations of the effect on their procedures and practices because they define the final objects ability to utilise the rights and powers assigned to them.

Note that this study, similarly to the stages approaches in policy analysis where policy process is divided into a series of stages such as initiation, estimation, selection, implementation, evaluation, termination[514], still adopts a legalistic, top-down bias, in which the

---

[512] Tala, *Lakien vaikutukset*, p. 70-74.

[513] Note that this is not a defining feature because the direct/indirect guidance provided by regulation and the issue of vertical packaging are analytically separate.

[514] DeLeon in The Stages Approach to the Policy Process, provides an introduction and a critical evaluation of the stages approach to the policy

focus is on the passage and implementation of a major piece of legislation. When the analysis in this study concentrates on individual instruments, this focus neglects the interaction of the implementation and evaluation of numerous pieces of regulatory instruments – none of them pre-eminent – within a given policy domain. The concentration on one individual regulatory instrument, e.g., the assumption of major piece of legislation, oversimplifies the usual process of multiple, interactive instruments involving numerous instruments at multiple levels (not even solely under government, but also wider).[515]

## 3.5 Classifications of regulators and their objects

In order to be able to analyse the regulatory capabilities of different instruments in its decentred form, a deeper understanding of who is regulating the issue of secure software development is needed. This depiction of the regulators, objects of regulation and other actors in the regulatory system concerning secure software development is made in order to be able to point out who is regulating and who is being regulated with specific instruments. This is no trivial task. If the regulators are desired to be identified, the roles they can play in different regulatory contexts has to be identified. The question of who is regulating, who is being regulated and what kind of an operational environment they create is important for the shape of the effects of regulation is going to have[516]. Different regulators have different resources and regulatory capacities accompanied, and thus different possibilities to influence behaviour.

The role of the regulated is not that obvious. It matters in relation to what the effects are going to be through the reactions of the objects

---

process.

[515] This part of the criticism towards the stages heuristic has been made explicit by Sabatier in The Need for Better Theories, p. 7.

[516] Tala, *Lakien vaikutukset*, p. 174 and more generally in chapters 8 and 9.

and the way they can be expected the react depends on the type of actor. The objects of regulation also have different informational resources available, their expertise in the issues varies, they have different levels of interests and they approach the issue from different perspectives.

Adopting a decentred understanding of regulation widens the range of actors that are in a position that enables them to intentionally influence the behaviour of others for the purpose of developing secure software and information systems (possible regulators)[517]. Regulators no longer are just state institutions but can also include a range of non-state institutions or actors (NSAs) and even individuals[518].

State institutions include a wide range of governmental actors on many levels. On national level there are parliaments, ministries,

---

[517] By limiting the study of regulation to intentional acts of directing behaviour, the regulators become distinguishable. When regulation is something intentionally used, there has to be recognisable actors having intentions. With a broader definition of regulation this would not be possible.

[518] This statement implies that legal authority is just one form of power. Although the state may have a near monopoly of such power, other key sources (information, expertise, financial resources, authority and legitimacy, organizational capacities and strategic position) are widely dispersed as pointed out by Black in Enrolling Actors in Regulatory Systems, p. 72-73. This dispersion of regulatory resources is one of the key arguments of decentring. The decentring thesis has emphasised that any of the regulatory tools may be used by the state, the community, associations, networks, or individual actors (including firms) with the obvious exception within constitutional democracy that the use of force and imprisonment are confined to the state, and that complex sets of relationships may exist between these actors. This makes the categorisation of different forms of self-regulation, more or less, analytically questionable; the same tools can be used by different actors and in many combinations between the actors may occur and there no longer is a basis for the black or white, state or self-regulation, debate. As Black notes in Mapping the Contours of Contemporary Financial Services Regulation, p. 3 and 9, state or self-regulation are just two examples of a far wider range of possible configuration of relationships and roles. However, this study does not go deep into these relationships and the regulators of secure software development (also their cooperative forms) are depicted only in the purpose of being able to identify the regulators and objects in relation to specific regulatory instruments used in the context of secure software development.

departments, regulatory agencies and governmental standard setting bodies; on supra-national level there is, e.g., the EU with its many parts. On international level bodies like the WTO, UN, OECD and NATO are important forms of cooperation between state institutions[519]. Also international governmental standard setting bodies, such as ISO are relevant. Naturally, there are many others, but recognising the different levels of state actors is sufficient because they are quite well studied and widely known. The usual assumption is that a hierarchically organised state or some hierarchical part of it successfully engages in ordering society.

Non-state actors (NSAs) are not all that well acknowledged. In the widest sense they include all those actors that are *not* representatives of states. They can basically be separated on the basis of their purpose; actors mainly aiming at producing financial wealth and driven by the goal of profit maximisation (economic actors, e.g. corporations, interest groups) and actors devoted to addressing public issues that do not aim to produce monetary profit for their members (societal actors e.g. voluntary organizations, social movements). Economic actors are less concerned with solving common problems or advancing a particular political agenda. At least four groups of NSAs can be identified; non-governmental organizations (NGOs), commercial pressure groups (business NGOs), corporations, and epistemic communities.

NGOs are non-profit, non-violent organised group of people, not established by governments, which are moreover not seeking government office[520]. It includes a wide range of actors from professional associations and international non-governmental standard

---

[519] These intergovernmental organisations could also be considered to be non-state actors (NSAs) due to their relative autonomy from states in making decisions and policies, given their expertise, formal authority, independent personnel and ties to non-governmental organisations. However, they are established by states, formally ruled by states and instrumental to state interests.

[520] For varying definitions of international non-governmental organisations, see Feld and Jordan, *International Organizations*, p. 21-24.

setting bodies to single issue organisations like environmental and human rights groups. Examples in the information security arena are the International Federation for Information Processing (IFIP)[521], non-governmental standard setting organisations such as the Internet Engineering Task Force (IETF)[522] and World Wide Web Consortium (W3C)[523], which typically are cooperative efforts between firms or individuals where organisational buyers (also individual buyers and other individuals such as researchers can participate) cooperate with sellers, and human rights groups like Privacy International[524] and Electronic Privacy Information Center (EPIC)[525].

Business NGOs are industrial interest (pressure) groups that, besides pursuing private goals, also seek to influence international politics. As such they are not-for-profit oriented, even though their members (corporations and business alliances) are not. Despite the similarity with NGOs they are separable due to their ideological and functional differences. Examples relevant for information security include commercial pressure groups like Business Software Alliance (BSA)[526] and Internet Alliance (IA)[527]. Even the Trusted Computing Group (TCG)[528] can be categorised under the business NGOs because it seeks after wide industry participation and wide adoption of the organisations specifications even though it is not-for-profit oriented.

---

[521] http://www.ifip.or.at/

[522] http://www.ietf.org/ Since IETF is a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet and is open to any interested individual, it could also be seen as an epistemic community.

[523] http://www.w3.org/

[524] http://www.privacyinternational.org/

[525] http://www.epic.org/

[526] http://www.bsa.org/

[527] http://www.internetalliance.org/

[528] https://www.trustedcomputinggroup.org/home

Generally, corporations themselves do not participate in politics. The industrial pressure groups (business NGOs) represent their interests. Nonetheless, today, individual firms and other market actors (notably actors who control key resources that firms need, i.e., "gatekeepers") increasingly further their interests in international arenas themselves. The difference is that corporations aim at maximising their profit while commercial pressure groups aim at furthering the private interest of their members (not necessarily directly related to profit maximisation). The main corporate NSAs are big international enterprises like IBM and Microsoft.

The last group, epistemic communities, is trans-national networks of experts with shared causal beliefs in certain policy issues. Examples relevant for information security are Computer Professionals for Social Responsibility (CPSR)[529].

A group of regulators that is difficult to fit into this category are gatekeepers[530]. Their position and power in the regulatory system is dependent on the situation and looks different on the basis of the perspective; they are not regulators because of their status or authority, but purely because of the resources they possess or otherwise control access to. Internet service providers, software component producers, different registrars and insurers are good examples. An example of gatekeepers in the secure software development are the reporters of vulnerabilities. In addition, the coordinators in the vulnerability reporting that act as intermediaries between the reporters and receivers of software vulnerability information can also be classified as gatekeepers[531].

Other relevant market actors are organisational buyers (organisational customers), individual customers or users in general

---

[529] http://www.cpsr.org

[530] Black in Enrolling Actors in Regulatory Systems, p. 70-71, uses the term gatekeeper to refer to actors who possess a key resource that a firm needs. Often they, instead the usual regulated firm, are made the targets of regulation.

[531] CERTs in their multiplicity of institutional background (some clearly governmental, some purely internal to commercial enterprises, etc.) are a good example.

(home/end user), manufacturers, developers and development partners as stakeholders, intermediaries who are selling or distributing the products manufactured by another seller in the market process (e.g. retailers, service provider or integrator), and other stakeholders like financiers, shareholders, subcontractors, suppliers and advisers (e.g., legal, risk and IT consultants, accountants). Then there are also evaluators of the products which can be external to the developing organisation (e.g., under Common Criteria) or even totally non-affiliated[532].

Even though the analysis of the regulators is focused on institutional actors and not on individuals because, in the aggregate, it is the institutions that have the best resources to influence behaviour, a class of individual actors has to be pointed out from the regulators perspective[533]. *Influential individuals* are advocates of a particular policy who, because of their reputation and institutional position, are at least instrumental and can be critical in ensuring that the particular policy is ultimately accepted[534]. These individuals are respected for their experience and advice. They play an important, though largely informal role in the regulatory process. In the young art of information security such persons are quite common[535]. Often

---

[532] For an overview of the major actors in a typical software vulnerability process, see Takanen et al., Agents of Responsibility in Software Vulnerability Process, p. 99-102.

[533] Although the final actors might be individuals (for example the developers of software code in relation to regulation that harnesses technology or social norms), they work within or are otherwise related to institutions. They are subject to the rules and norms of these institutions, thus attenuating individual preferences or desires. This does not mean individuals are irrelevant. Moreover, the institutional values and preferences are a composite reflection of the individuals compromising these institutions. And the effects of regulation are always about the actions of individuals.

[534] This term is also taken from Black, Enrolling Actors in Regulatory Systems, p. 71-72.

[535] Many of the recent changes to improve security and privacy for the Internet are the direct result of motivated individuals, rather than policymakers in government.

also individuals inside firms, such as executives, have possibilities to influence behaviour just with they example and sayings. One only needs to think about a few executives in major corporations.

These actors are, at the same time, both regulators and being regulated at different levels. They can and do cooperate in different ways, to the extent that different networks and hybrids combining both state institutions and non-state institutions have been recognised as qualitatively different forms of regulators. Thus, in addition to the traditional regulatory models of hierarchies (typical to state institutions and many organisational form NSAs, even though there is an internal move towards network and hybrid forms in them), contracts (bi- or multilateral) and organisation (associations), regulation often occurs via network-like interlinkages of different actors[536]. Traditionally the emphasis has been on hierarchy as an instrument of control, but nowadays there is a shift towards other structures for control. But as expressed by Scott, this change may be one of thinking rather than underlying mechanisms, since it has long been clear that the period of organised capitalism is characterised by a mixture of state, market, community, and technology within social control processes[537].

## 3.6 A methodological aside – on the role of law in regulation

Even though this study is focused on regulation and not just on the law, the role of law in regulation is naturally of interest already because laws are the major ways in which western democracies structure

---

[536] Even though networks have their own important characteristics as regulators (involve a variety of actors each pursuing their own goals, between whom there are relatively stable sets of inter-relationships, and who are dependent on one another for resources) the seldom act alone. Typically they combine contracts and associations. The analysis of the regulators could be taken much further, but is not beneficial in considering the different tools of regulation and especially law's role in regulation. Any of the tools maybe used by the different actors, with the self-evident limitation that the use of force and imprisonment are confined to the state.

[537] Scott, Regulation in the Age of Governance, p. 18.

themselves and they still are the major form of implementation of policies and agreements at supra-national and international levels. The role of law is taken further in the contextual considerations of each specific instrument. Despite of this, a general description of the role of the law is elicited.

My purpose is not to discuss about a generally applicable concept of law in regulation. Neither do I try to make grand generalisations about the problems with the law and policy relationship. As Terence Daintith points out in his introductory essay to a larger collection of studies on the instruments used in the implementation of economic policy and the role of law as an instrument for implementation, national difference in the discussions between the problems of the law and/or policy relationship and the concerns about instrumental approach to law suggests the futility of such arguments[538]. The way the problem is perceived depends too much on where you come from. In here the interest lies only on the ways the conception of the law affects the way regulation is seen to influence behaviour.

There is no clear answer to the problem of the relationship between law and regulation. At minimum, the answer is dependent on our understanding of both; the way both the 'law' and 'regulation' are defined is the first problem to overcome[539]. Above, regulation was defined along the decentring thesis and was seen as an intentional action. The question of what the 'law' is has troubled lawyers for ages and there is no purpose in going too deep into this discussion[540]. Suffice to say that the centred and decentred (pluralistic) conceptions of law differ similarly to the conceptions of regulation. While there is at least a sort of an agreement what law is in its centred form[541],

---

[538] Daintith, Law as Policy Instrument, p. 4-5.

[539] This argument is made by Black, e.g., in Law and Regulation, p. 54, footnote 78.

[540] For a recent discussion of what 'law' is from the point of view of 'decentring' or 'pluralism' see Tamanaha, *A General Jurisprudence of Law and Society*; idem., *Realistic Socio-Legal Theory*; Twining, A Post-Westphalian Conception of Law, p. 199-259.

[541] The classic is Hart, *The Concept of Law*.

at the same time legal pluralists agree that 'law' does not solely emanate from the state. What it is and how it is distinguished from other forms of norm-based social ordering is still contested, as Black points out together with extensive references[542]. What this means is that the relationship between law and regulation is as complex and shifting as the conceptualisations ascribed to each.[543]

The instrumental rationality behind most regulation (including much public law) is quite different to the ideal of legal reasoning presented in relation to much of civil law (especially when based on civil codes) or the similar construction of judge made law[544]. The ideal is internal coherence and autonomy from operational or outcome considerations. As Ugo Mattei notes in developing comparative law and economics as a discipline, this unitary theory of the law is not just the common and unchallenged idea of the two major legal theories in Western jurisprudence (naturalism and positivism) but also still widely shared by the lawyers in both civil and common law systems[545]. With regulation, by contrast, problems in widely diverse domains tend to entail quite context specific goals and techniques. Instead of seeking

---

[542] Black, Critical Reflections on Regulation, p. 23.

[543] In agreeing that law or regulation does not solely emanate from state the pluralistic and the decentred views leave basically only two possible options for the conception of 'law' and 'regulation': to abandon any attempt to hold on to a single coherent conception or to attempt to construct a minimalist core concept. Twining, A Post-Westphalian Conception of Law, p. 206. My purpose is not to argue for or against either of the strategies; the only thing I am worried about in this the way the chosen strategy affects our understanding of 'regulation' and 'law'.

[544] As Mattei notes in relation to the usability of capture theory in different legal systems, "[c]omparative law literature makes it quite clear that the function served by civil codes is analogous to that served in America by the common law rather than by statutes" (Mattei, *Comparative Law and Economics*, p. 76). Civilian codes are much more products of the legal culture than of the legislative process. Note, however, that growth of the civil law – and all of public law – happens much more often by special legislation outside of the code, which can be compared to statutes in the common law, or by case law rather than by tampering with the language of the code.

[545] Mattei, *Comparative Law and Economics*, p. 102-103.

coherence of principles, the policy maker or regulationist asks which rules (or other techniques) work best[546].

While considering regulation to be 'more than law' in that law is only one of the techniques or instruments that may or may not be involved in the practice of regulation[547], regulationists are engaged in the same practice as socio-legal scholars. They look at the 'law' (however defined) as an outsider of the legal arena. This external view treats law as one component of social ordering, to be studied for the ideas it embodies and the effects it produces[548].

There is, however, a critical difference between socio-legalists and regulationists; even where each asks the same questions, they do so for different reasons. Regulationists are not (and by implication socio-legalists are) interested in how law sees itself, explains itself or legitimises itself; "…not only is regulation *not just* law in that it extends well beyond courts and legal instruments, regulationists are *just not* concerned with law in that they are not concerned with whether or not law is correct in seeing itself as characterised by unity, coherence or particular modes of reasoning, or explaining itself in these or any other terms"[549]. Thus, socio-legalists focus on, and the sociology of law as a discipline or a methodology is devoted to, an inquiry into

---

[546] This distinction is often made in regulatory studies. See, e.g., Smith, What is Regulation?, p. 39; Tala, *Lakien vaikutukset*, p. 79; Niemivuo, *Kansallinen lainvalmistelu*, p. 10-13. In a more generalised version this develops into the competitive relationship among sources of law as proposed, e.g., by Mattei in *Comparative Law and Economics*, p. 101-122.

[547] This implies a centred conception of law; it emanates from the state.

[548] This dichotomy of legal scholarship into the internal and external perspective is the basic approach of textbooks in legal theory. See, e.g., Rubin, Legal Scholarship, p. 562, and Aarnio, *Laintulkinnan teoria*, p. 54-56. This dichotomy was originally developed and made famous by H.L.A. Hart who elaborated it into a method of understanding legal and other human institutions by reference to their meaning from insider's or an internal point of view (McCormick, *H.L.A. Hart*, preface).

[549] Black, Critical Reflections on Regulation, p. 26.

the internal structure and meaning of the legal system and in doing so become a part of the legal scholarship[550].

Regulationists, on the other hand, dissociate from legal scholarship by refusing to give any significance to the internal view of law in the regulatory scholarship. In terms of Black "…regulation, however it is defined, has no claim to be unified, or coherent, or marked with its one style of reason or argument, or based on consistent values or principles. Nor does it invoke or lay claim to any mystique or even legitimacy. … Regulationists, and others, do not expect regulation to be internally rational or consistent; it might be, more likely it will not. But no significance attaches to either conclusion (though regulationists might worry about the impact of such inconsistencies on regulation's effectiveness). In contrast, such internal unity and consistency is central to many understandings of law."[551]

However, the instrumental approach is unavoidable even in the internal views of law. Even the acceptance of the narrowest conception of regulation, as a form of a legal rule or regulatory ('public') law where it essentially is 'less than law', necessitates an instrumentalist view of the law. Most legal scholars at least implicitly accept that legal rules are adopted to promote some goal, be it equality, justice, fairness, or efficiency[552]. This, in turn, poses a different set of questions about the legal system; those of effectiveness rather than coherency. It also demands a different task for law and legal theory. This is troubling for legal scholars (sharing the internal view of law); the pluralistic (decentred) views are fragmenting law internally and making it difficult to hold the coherence thesis.

---

[550] See, e.g., Rubin, Legal Scholarship, p. 562.

[551] Black, Critical Reflections on Regulation, p. 25.

[552] Legal reasoning is not and, as Collins in Regulating Contract Law, p. 18-19, argues in relation to goal oriented regulation of contracts and contract law, cannot be purely about coherence considerations. When it is accepted that there is no coherent scheme of individual rights embedded in the social and political institutions of a society according to which legal reasoning could proceed, there necessarily is a need to employ also consequential or policy arguments in order to reach determinate results.

Regulatory theory, similar to the social scientific analysis of the law, once again questions the role of law in the steering of societies. However, with decentred regulation the potential for fragmentation is even greater. Decentred regulation is likely to adopt institutions and practices which are not found at all in traditional accounts of law[553]. At this point the borders between regulation, law and social control become quite permeable if they do not disintegrate entirely as pointed out by Smith[554].

I agree with those regulationists that consider regulation as 'more than law'[555]. But their view of law is limited; the extreme instrumental view prohibits the acceptance of law's other functions. While they dissociate from legal scholarship by refusing to give any significance to the internal view of the law in the regulatory scholarship they, at the same time, dismiss the value of the legal system. They get, at best,

---

[553] However, they may be encompassed by legal pluralism.

[554] Smith, What is Regulation?, p. 39-40. However, this does not mean that it is regulationists that threaten the unitary approach to law (the internal view of law as a coherent system having value as such). The issue concerns only whether or not regulatory studies can be considered as legal scholarship or is it occupying an intermediate position as an interdisciplinary approach. The instrumental approach to law together with the decline of state as the sole authority of law (destruction of the traditional hierarchical system of law) and the pluralism of legal sources is much more troubling for legal theory. Even though I accept, to some degree, the point of view of a regulationist and an instrumentalist, these considerations are not taken further in this study; despite of their importance for legal scholarship.

[555] The notion that regulation is 'more than law' is strongest when a decentred conception of regulation is invoked and set in relation to a centred conception of law. Black, Critical Reflections on Regulation, p. 25. According to Black, idem, p. 25, lawyers see the lack of coherence requirement in regulation meaning that regulation is 'less than law'; it is a differentiating factor between law and regulation. But from a regulationist point of view the situation is reversed. The coherence requirement and the internal view to law restrict the use of 'law' (legal norms and legal order, i.e., the collection of legal norms in force at the specific moment) as a regulatory tool; 'law' is just one way of regulating. This is acknowledged also by Black, idem, p. 25-6, by pointing out that when regulationists look at legally based regulation they are not bothered by the internal view of law; regulation is something else and something more than law, however it is defined.

only an inadequate understanding of the meaning that the law possesses for its members (the actors of law; legislators, judges, legal scholars). Because such meanings are components of a comprehensive life world, they cannot be fully understood unless the observer participates on the same terms as the members[556]. By dismissing legal scholarship regulationists do not see the way the internal view of law affects the regulators and other constrains (state institutions, non-state institutions, market forces, social forces, technology).

As pointed out by Alan Hunt the law should not be seen as a mere medium of regulatory intervention[557]. Law is not just one of the techniques or instruments that may or may not be involved in the practice of regulation; it is not just legal norms or legislation. The legal order and the legal system is an internally coherent construction that bears significance as such and provides information not attainable by viewers external to its operation. When viewing law internally, as legal scholarship does with its three basic methodologies (descriptive scholarship, prescriptive scholarship, jurisprudence), it is a set of significant normative statements that are intended to comprise a meaningful system; as such, its provisions should be described in detail and evaluated according to their moral or social value. Law can be considered as an expression of grounding moral values, religious beliefs, well-established custom, universal principles of human rights or conceptions of justice that exist despite of political decisions or changing objectives. What is common to all of these is the unitary approach (the coherence of legal system); law is considered as a coherent system that is based on consistent values and principles.[558]

As Kaarlo Tuori points out in discussing the positions of the participant and the observer in the analysis of the law, approaching

---

[556] This is the classic argument of legal theory. See, e.g., Rubin, Legal Scholarship, p. 563; Aarnio, *Laintulkinnan teoria*, p. 54-56.

[557] Hunt, The Politics of Law and the Law of Politics, p. 68. As should be clear, neither should the "forces" that are harnessed for the purposes of regulation, i.e., social norms, markets and technology, be seen in this way.

[558] Note that this deviation from the strict instrumental view of law still implies a centred conception of law (emanates from the state).

law as social practices may lead the social scientist from her position of an outside observer into neglecting the inner mechanisms of the legal field through which the inputs from the political, formative and cultural symbolic dimensions are filtered. The social scientist maybe sensitive to the symbolic normative side of the law only as one of the motivational or communicational factors of legal practices, but less so to the specific transformative legal work of interpretation and systematisation producing and reproducing the law and to the power relations between legal professionals linked to this work.[559]

As visible in the classification of regulatory instruments, much of the role of the law is part of the prescriptive rules category. This is the strict understanding of the state centred form of the law as expressed in regulative law. Regulatory scholarship tends to restrict its analysis of the law as a regulatory instrument to this perspective[560]. At a cursory reading regulatory scholarship easily suggests that the role of law is non-existent in regulation in general. However, this is not the whole picture of the law. Not only duty imposing legal norms are relevant in regulation. The law is in the background of much of regulation and of most types of regulatory instruments; and not just when the government is using these instruments.

---

[559] Tuori, Law, Power and Critique, p. 21.

[560] This might be due to the economic analysis of the law, one of the main scientific disciplines behind the new construct of regulation as a separate field of scientific inquiry, mainly concentrating on legal sanctions as influencing factors. As pointed out by Hugh Collins in *Regulating Contracts*, p. 120-121, even if the Chicago school type economic analysis of the law (hypothetical cost-benefit analysis) would be widened to include also non-legal sanctions, it is still flawed in its explanation of contractual relations and more widely as the way regulation influences behaviour. It concentrates only on punishment and does not analyse the law as a signal of credibility of the commitment etc., which for Collins means that institutional economic would provide a better answer. When the Chicago school type of economic analysis of the law adopts legal sanctions as its central focus, and widely ignores non-legal sanctions, their relationship is easily perceived as exclusive. When the forceful impact of non-legal sanctions is compared to the relatively insignificant legal sanction in the regulation of the markets, the law is easily seen to play a minor role. At the same time their interrelationship is easily ignored.

A central part of the understanding of regulation (at least for a legal scholar) is the way law is used to affect other regulatory means[561]. The conception of the vertical packaging of instruments is essential when considering the role of the law in the decentred process of regulation. Law, especially in its indirect sense, is at the top of vertical packages. Legal norms are the main instruments still used to regulate the influence mediating actors at the lower levels of packages and to govern the instruments they use. Without that concept and the accompanied regulatory strategies the role of law would be close to inexistence when the state is taken away from the centre of regulatory power. This legal underpinning for indirect control over other ordering systems is the basis for consensus on the role of law in regulation as suggested in the studies of the post-regulatory state[562]. Thus, state institutions (at multiple levels) still retain legal control at least in some residual form; e.g., goals and ends are ultimately set and determined, even though not necessarily directly, by state-centric actors[563].

---

[561] Black in Critical Reflections on Regulation, p. 22-27, considers the issue widely. She studies the relationship between law and regulation at the conceptual level and considers both the possibilities for fragmentation and the problem it faces especially to our understanding of law, but does not go deeper into their interaction or to the influences of law to the different decentred regulatory (in the wide sense) activities. Of course there are interactions between the other forms of regulation too, as noted above in relation to vertical packaging, but in here the interest is in how law is used to affects the other instruments.

[562] See, e.g., Scott, Regulation in the Age of Governance, p. 21; Teubner, After Legal Instrumentalism, p. 307; Black, Decentring Regulation, p. 104-105.

[563] Scott, Regulation in the Age of Governance, p. 21; Lessig, The New Chicago School, p. 666-667. Naturally if other tools of regulation exist or can be employed, they might be preferable to regulatory law because they may be more effective or more efficient. In addition, if the incentives for acting within competitive or community control structures are sufficiently aligned with conceptions of public interest, there might not even be a need for the state law underpinning. However, a strategy to delegate management of a certain issue or a regime to non-state actors can also be a conscious state choice. In this way, when it is done in a way that not only takes notice of the public interest but also tries to guarantee its realisation, state-actors can be seen to retain

Much of the study of the other ordering systems ("forces") themselves (the market, social norms, technology, and even processes[564]) that regulation as an intentional act is considered to be harnessing in this study, argue against the dominance and centrality of law; that these other ordering systems displace the significance of the law[565]. From the decentred perspective the law is easily seen to be displaced as the sole or even primary regulator of behaviour. But the argument, inline with the Lessig's New Chicago School argument, is that instead of displacing the law they are subject to the law and regulation more generally[566]. The ways the law affects behaviour just need to be understood more widely. Thus, rather than diminishing the role of law, these alternatives suggest a wider range of regulatory means; different instruments can be used by a multiplicity of regulators to harness these "forces" for their own purposes[567]. By combining

---

control even in the global governance.

[564] As Teubner argues, among others, the way the law regulates has changed from regulatory law to reflexive, procedural or post-regulatory law. Where regulatory law set substantive standards, reflexive and procedural law sets procedures.

[565] These types of studies in the area of economic analysis of the law are playfully referred to as the 'Old Chicago School' by Lessig in The New Chicago School, p. 665-666.

[566] In terms of Lessig "… in the view of the new school, law not only regulates behaviour directly, but law also regulates behaviour *indirectly,* by regulating these other modalities of regulation directly." (Lessig, The New Chicago School, p. 666). While the old and new 'Chicago schools', as Lessig playfully calls them, are united in their view of law as just one of the regulatory mechanisms, the new school does not see these alternatives as displacing law but, instead, as subject to law. Lessig, The New Chicago School, p. 661-667. Note that Lessig does not separate between vertical packaging of instruments (the perspective of the regulator) and the direct/indirect guidance provided by regulation (from the perspective of the regulated). Even though they are highly related issues, they still a separable and need to be distinguished. He seems to be primarily speaking about vertical packaging on the basis of the examples he uses.

[567] When the government uses these instruments, they typically are enacted in laws. Regulators, not even the government by using the law, cannot harness the other ordering systems perfectly, completely and especially not always easily or in obvious ways, but it can be and is being done.

the New Chicago School and decentring theses, it can be stated that while there is a reduction in the regulatory role of the state, it does not mean that some other ordering system gets dominance.

The importance of the legal norms is visible especially in relation to the power conferring norm type. These are what constitute new institutional structures or strengthen existing ones as is the case of the market economy needing at least private law to operate even though it can exists in spite of the legal system. And the ability to regulate is essentially law based not just in the case of the governmental actors (especially in constitutional democracies) but also in the case of private regulators like that of the firm[568].

In sum, the role of law depends on how it is understood. If considered, as in here, to be state centred, its role is diminished but not marginal. It is still a major part of the prescriptive rule category in its regulatory sense and in the background of most of other instruments especially when used by the state. The indirect strategies are essential for the understanding of the role of the law. Even though the regulation is decentred and the legal sources are plural, the law is a central element shaping the other regulatory instrument and being shaped by them. The law is at the same time both regulating behaviour and being regulated by state and non-state actors using different instruments. [569]

---

[568] Market actors can use private law powers to create more limited private rights, for example to demand collaterals, or require certain disclosures, assuming they have the market power to do so (and the law will support their efforts). Black, Mapping the Contours of Contemporary Financial Services Regulation, p. 4. With powers conferred by private law even a private citizen becomes a private legislator. By the possession of the powers she is made competent to determine the course of the law within the sphere of his contracts, trusts, wills, and other structures of rights and duties which she is enabled to build as has been pointed out in legal theory already by Hart in *The Concept of Law*, p. 40-41.

[569] The case where the law is understood in a plural sense, which is a subject of heated debate, is not taken further here because the relationship with regulation becomes too blurred.

*Transition*. It must be kept in mind that there is no reliable knowledge of the ways regulation affects behaviour in practice. Current knowledge is not comprehensive. Regulation together with other motivational factors determine in complex and still largely unknown ways the individual choices to comply with regulation[570]. Armed with a preliminary understanding of the types of instruments (classified as formalised rules, market harnessing and economic, informative and pedagogic, technology-harnessing, and process-based) and of their normative guidance mechanisms that can be intentionally used to influence behaviour, we can now, however, turn into studying the specific instruments of secure software and information systems development regulation, their types and the capabilities to influence behaviour.

[570] As pointed out, e.g., by Tala in *Lakien vaikutukset*, p. 323.

# 4 Harnessing social norms: disclosure of vulnerability information

As discussed above, economic theory assumes that in perfect markets vendors have an incentive to disclose all the relevant information, also the negative information that might hamper the sales of the product. However, due to the imperfections in the COTS software markets, vendors do not have high incentives to disclose the negative features such as vulnerability of their software voluntarily.

By summarising the above discussion on the inadequacies and asymmetries in the distribution of security related information, it can be stated that in the situation where the customers have no easy and cost-effective means to differentiate between vulnerable and secure software and thus to verify the claims about quality and security made by the vendors and, in addition, where the customers have limited means to calculate the losses from vulnerabilities and thus tend to underestimate the risks and not to demand more security, the vendors have high incentives to behave opportunistically and to take advantage of the informational misbalance[571]. Because of the inadequate incentives for informing the consumers about the risk on the part of vendors and the limited capacity of the consumers to process the available information, unregulated market transactions will not result

[571] See heading 2.8 above. A central feature of opportunistic behaviour is that a party has a motive not to disclose certain information, or to make false or misleading claims, if the difference in the level of knowledge between contracting parties brings advantage. Note that not every vendor is behaving opportunistically, but the mere number of those who do, together with the impossibility to separate them from the non-opportunistic ones, and the extent to which such motivation exist in market-driven development, makes the issue important.

in optimum care by the firms and optimum consumption by the consumers, when the latter are imperfectly informed about the risk.

There is a need for some way out of the loop where vendors do not have sufficient incentive to produce secure software, especially due to lack of customer demand, and the customers, even if their acceptance of the defects might be diminishing, as it currently seems to be doing, thus creating more demand, cannot separate high quality and secure software from the vulnerable ones and thus make market rational choices and to demand more secure software effectively. The above analysis of informational asymmetries in the market for secure software made explicit that the market forces themselves have certain mechanism to correct the informational market failures by communicating quality and security convincingly[572]. However, they are only partial solutions accompanied with several inadequacies.

In addition to the non-applicability of the self-corrective measures to credence characteristics like the security related vulnerabilities that require sophisticated technical assistance even for information system developers in order to be determined, and the several limitations in their applicability to experience characteristics, the inadequacies stem from the basis that the information coming from the parties that benefit financially as a result of the reaction to the information (e.g., from a decision to procure the advertised product) is biased and the provider of the information *can* have a motive to hide certain negative characteristics, or even mislead the customer. Customers need to find more objective information than provided by those providing also the commodity that the information concerns, especially of credence features like security related vulnerabilities[573]. The role of informative instruments that decrease the risk of deception and withholding of

---

[572] See heading 2.8 above.

[573] Since credence characteristics of goods are qualities which the consumer rarely discovers by her own personal search, in the absence of independent additional information the consumer cannot discover whether an operation in the software development was justified and the correct procedure for the fixing vulnerabilities applied.

negative information is crucial under these conditions[574].

Regulatory literature recognises two general types of informative instruments[575]. The first involves *direct supply* of information to the addressees by the regulator (by its own organisation). The second involves the *use of intermediaries* to transmit the message to the public. In the latter case the regulator itself does not possess the information; it requires or enables other actors to generate or share the needed information. Basically sets in motion a process of information collection or learning.

*Enabling* involves paying intermediaries to collect, package, and disseminate the needed information and thus involves vertical packaging with economic or market-harnessing instruments. An example can be found from the funding of basic education of software engineers and of others responsible for software development in issues of secure development. *Requiring* involves the use of prescriptive rules to promote dissemination of information. An obvious example is the case of mandatory disclosure of certain product information (like quality or safety) and the control (prohibition of the supply) of false or misleading information[576].

---

[574] Note that despite of what motivates vendors to dismiss the security related vulnerabilities, i.e. whether it is the fear of negative publicity or the lack of understanding of the nature of the problem and risks involved in a situation where vulnerabilities are easily considered to be un-exploitable or their exploitation considered to be very unlikely, the disclosure rules bring relief for the customers. The latter motivation is altered by making the disclosure mandatory and the former by informing both the vendors and their customers of the importance of vulnerability issue.

[575] They have especially been identified in the new tools approach to governance studies, for example, by Weiss in Public Information, p. 218-219, and Vedung and van der Doelen in The Sermon, p. 106-107. Also regulatory studies, such as Baldwin and Cave, *Understanding Regulation*, p. 49, and Ogus, *Regulation*, p. 121, use this differentiation. Note them being in line with the vertical packaging and increased use of third-party arguments already made.

[576] As Collins notes in *Regulating Contracts*, p. 292, the requirements for the manufacturers and retailers to supply detailed and reliable information about the hidden costs and risks to users of the product is the most obvious and widely used solution to the problem of quality and security created by informational asymmetries. In other words, regulation involves extensive duties

Informative instruments are typically the first ones to which regulators turn to when deciding to alter behaviour into certain direction. The already discussed argument made in regulatory studies of the escalation of instruments (starting from less coercive instruments) and their chronological sequencing is clearly visible also in information security. Informative instruments have been the first areas of focus, for example, when governments have implemented or are implementing the OECD Security Guidelines[577].

In this study, only one informative instrument is analysed. It is that of *vulnerability reporting* by external parties (external to the vendor that developed the specific software where the vulnerability is found, e.g., individual or organisational customers) both to the vendors themselves and to the public[578]. This is a form of non-governmental disclosure regulation which importance has unfolded in information security in the past ten or so years. It carries the characteristics of both

---

of disclosure about such matters as safety risks, costs of repairs etc.

[577] OECD *Guidelines for the Security of Information Systems and Networks*, Recommendation of the OECD Council at its 1037th Session on 25 July 2002. Of the 21 OECD member states that answered the survey on the implementation of the OECD guidelines for the security of information systems and networks all, except one, report initiatives that aim at awareness rising in 2004. Conference-type events, Web sites and publications are most frequently mentioned. OECD, *Summary of responses to the survey on the implementation of the OECD guidelines for the security of information systems and networks*, p. 5. One of the main findings of the subsequent OECD report analysing the responses, *The Promotion of a Culture of Security for Information Systems and Networks in OECD Countries*, p. 3, is that awareness raising and education initiatives still receive a high degree of attention.

[578] Vulnerability reporting has been defined, in the first empirical analysis of the communication in the vulnerability reporting process by Tiina Havana and Juha Röning, Communication in the Software Vulnerability Process, p. 2, as a process that refers to the communication in which the knowledge of a vulnerability is transmitted to the persons or organisations responsible for fixing or distributing the knowledge of the vulnerability further to other relevant parties, such as customers. A good resource into the various aspects of, approaches to, and interests in vulnerability disclosure debate are the pages of the Oulu University Secure Programming Group (OUSPG) "Vulnerability disclosure publications and discussion tracking" at http://www.ee.oulu.fi/ research/ ouspg/ sage/disclosure-tracking/ [22.2.2006].

general types of informative instruments; both the direct supply of information and the use of intermediaries.

Several other, especially governmental types of disclosure rules are omitted. A separate analysis of the rules controlling the supply of false or misleading information, the negative duty not to misinform the other party to a contract (not to lie), is not conducted because these legal rules (e.g., basic provisions on fraud and misrepresentation and its many variants for example in marketing regulation) are already a prerequisite in the economic analysis of the producers incentives to reveal negative attributes of products[579]. A variant of the disclosure rules, the consumer protection oriented obligations to provide information under the threat of sanctions (form of mandatory disclosure)[580], is non-existent in the context of the information concerning the vulnerabilities of software products and omitted for this reason. The third type of disclosure rules that in the traditional state-centred approach to regulation involves only the direct supply of information by a scrutinizing regulatory agency or a governmental official is also non-existent in the software industry.[581]

Vulnerability reporting is a disputed and controversial issue in current information security discussion. Vulnerability management in general and reporting especial have become hot topics quite recently

---

[579] See e.g., Beales, Craswell and Salop, The Efficient Regulation of Consumer Information, p. 502. This is also the starting point in the above analysis (heading 2.8) of the informational market failures in information security.

[580] An example can be found from the obligations to disclose safety information under product safety regulations such as the article 5 of the Directive 2001/95/EC of the European Parliament and of the Council of 3 December 2001 on general product safety, Official Journal L 011, 15/01/2002, p. 4 – 17.

[581] When accepting the regulatory role of private law, i.e., accepting that private law even in its non-mandatory form is not just facilitative, the disclosure rules become wider. This makes it possible to see that different liability rules under doctrines in contract or damages (tort) law and even general marketing rules create obligations to inform and to disclose vulnerabilities. However, since the informative character of these rules is only one of the ways they can influence behaviour, it is analysed in the prescriptive rule section together with its other influence mechanisms in chapter 5.

and the practices and procedures are just emerging and consensus still being sought[582]. Because the regulatory function is not plainly visible and has not been accepted by all, there is a need for further justifications.

The regulatory function of vulnerability reporting is best visible, and has been most forcefully argued, in the full (and immediate) disclosure movement that explicitly attempts to alter the behaviour of major commercial vendors[583]. Vulnerability reporting can be seen as a regulatory action since the objective is to alter the behaviour of both the vendors (to correct the vulnerability and to issue a patch) and the users (to take protective measures, e.g., by installing patches or by reconfiguring the system to minimise the effects of the vulnerability) of the specific software in order to achieve improved information security. I.e., the process of vulnerability reporting involves a sustained and focused attempt to alter the behaviour of two different stakeholders with the intention to improve the security of software.

Even though the approaches to the reporting of vulnerabilities still vary significantly even among the external parties[584], a rough consensus among those involved in the reporting process is that disclosure (in some form) is needed for the improvement of the security of software. The customers and the public need to know

---

[582] This makes the study of the possibilities of this type of regulation to affect secure software development difficult, but not impossible. However, due to the rapidly evolving discussion the considerations of the possible influence on secure software development can, at best, be only tentative.

[583] The unresponsiveness of software vendors is one of the reasons the issue of full disclosure and the threat of disclosure of vulnerabilities to the wider public in general has been considered to become issues in the first place.

[584] Others favour full, public and immediate disclosure whereas others partial and limited public disclosure after a predefined time, or something in the between. This is similar to the differences between the opinions of the reporters and receivers of the reports, but less variable than between all the stakeholders of the reporting process (the discovers of the vulnerabilities, the vendors of the vulnerable software, the users of the software, and the coordinators of the vulnerability reports) as reported by Havana and Röning in Communication in the Software Vulnerability Process, p. 9.

about vulnerabilities and the available remedies, and the disclosure is needed for reducing risks to information systems and for stopping malicious activity; disagreement centres on how, when, and to whom to disclose[585]. Thus, the enhancement of the security of software in general is the expressed common objective of the different vulnerability reporting schemes.

Even though both the importance and the media attention of full and immediate disclosure may have diminished due to it partly accomplishing its objectives (i.e., software vendors take the reports by external agents more seriously and patch the disclosed vulnerabilities) and being replaced by more mature versions like responsible disclosure[586], the regulatory value of vulnerability disclosure has not vanished. It has just been reshaped.

Whereas vulnerability reporting originally was only about the action of the discoverers/reporters of vulnerabilities, now it can be seen as a type of self-regulation where the various stakeholders of vulnerability reporting adopt common guidelines amongst themselves and for themselves. Not only has vulnerability reporting gained a central place in software quality and security development in COTS software development companies in the sense that also proprietary software vendors increasingly rely on the disclosure process in their quality

---

[585] This consensus becomes explicit in the first empirical analysis of the communication in the vulnerability reporting process conducted by Tiina Havana in 2002 and reported in her M.A. thesis for the Department of Communication at the University of Jyväskylä in Finland, *Communication in the Software Vulnerability Reporting Process*, p. 45-46 and 56-58. In the final report of the Vulnerability Disclosure Working Group of the National Infrastructure Advisory Council, which was charged with advising the President of the U.S. on information system security issues important to preserving the integrity of the nation's critical infrastructure, Chambers and Thompson, *Vulnerability Disclosure Framework*, p. 7-8, 16, and 26, conclude that discoverers and vendors share the primary goal of improving the security of software used in critical processes. Note that the report also refers to the research of Tiina Havana and the Oulu University Secure Programming Group (OUSPG).

[586] The consensus among the stakeholders in the vulnerability disclosure process seems, as discussed below, to be in favour of disclosure of certain amount of information after predetermined time. However, to the remaining proponents of full and immediate disclosure it still appears as the only way to force vendors to take vulnerabilities seriously.

and security testing (they have become amazingly responsive to vulnerability reports)[587], but also common policies and guidelines for vulnerability reporting are under development[588].

The importance of the regulatory function of vulnerability disclosure is acknowledged even by governments. The wide establishment of even governmental vulnerability coordination centres together with policies and guidelines for the reporting process testify this. As the OECD in its follow-up study on the implementation of its 2002 Guidelines for the Security of Information Systems and Networks in member states emphasises, most countries responding to the survey support the establishment and use of CERT-like sites. Some already report the establishment of CERTs for government or CERT initiatives targeting SMEs and/or the general public[589]. With

---

[587] This change of policy among the software vendors is no surprise in market-driven development since the vendors can lower the costs of testing by harnessing users for that purpose. Thus, instead of pretending nothing is wrong or that their customers do not care about security, the vendors have become more and more responsive to vulnerability reports and are developing more formal and mature processes for it. This change in the attitude of the vendors is partly due to the regulatory action taken by the reporters and it is part of the analysis done below.

[588] In addition to the discoverers/reporters (individuals or organisations that find the vulnerabilities), explicit policies for vulnerability reporting have been presented also by the vendors (parties that develop or maintain software products that may be vulnerable) and the coordinators of the communication. Whereas policies and guidelines for vulnerability reporting have originally been developed by the different actors separately, currently the trend is towards common guidelines for all or almost all actors. Most prevalent of current proposals are the OIS (Organization for Internet Safety) *Guidelines for Security Vulnerability Reporting and Response* and the NIAC (National Infrastructure Advisory Council of the U.S.) Guidelines that are presented in the final report of its Vulnerability Disclosure Working Group drafted by Chambers and Thompson, *Vulnerability Disclosure Framework*. For a deeper discussion, see below heading 4.3.3.

[589] OECD, *Summary of responses to the survey on the implementation of the OECD guidelines for the security of information systems and networks*, p. 13. As the subsequent OECD report, *The Promotion of a Culture of Security for Information Systems and Networks in OECD Countries*, p. 3, states while analysing the responses as one of its main findings, international cooperation is consolidated in this area.

such governmental support we may even speak of some kind of a joint regulation between government and the private sector[590].

> Even though the OECD speaks of CERT organisations, it is more appropriate to speak of Computer Security Incidence Response Teams (CSIRT). CERT together with CERT/CC is a registered service mark (U.S. Patent and Trademark Office) of the Carnegie Mellon University[591]. CERT/CC has for its part helped to establish

---

[590] With joint regulation I refer to the coexistence and combination of endogenous 'self-regulation', i.e., the common guidelines adopted by the economic actors, the social partners, non-governmental organisations or associations amongst themselves and for themselves, and exogenous governmental regulation. When the government has a deeper and a more direct involvement, e.g., in the form of setting the essential legal framework and then monitoring the outcome or even validating the more detailed rules filled in by the stakeholders, reference is typically made to co-regulation. For a discussion of the role of self-regulation and co-regulation in the European regulatory policy, see Linda Senden, Soft Law, Self-Regulation and Co-Regulation in European Law. As Philip Eijlander argues in Possibilities and Constraints in the Use of Self-Regulation and Co-Regulation in Legislative Policy, heading 5 "Concluding remarks", first four paragraphs, both the concepts and the use of the types of regulation differ in European and national levels due to the different functions of national and European legislation. In the end, both self-regulation and co-regulation are examples of a far wider range of possible configuration of relationships and roles of different actors in a regulatory setting and the separation between state and self-regulation is analytically questionable in a deeper regulatory analysis. For references and a wider discussion, see above footnote 518.

[591] The CERT Coordination Center (CERT/CC), the first organisation to coordinate communication among experts during information security emergencies, grew from a small computer security incident response team formed at the Software Engineering Institute at Carnegie Mellon University (Pittsburgh, Pennsylvania, USA) established by DARPA (Defence Advanced Research Projects Agency) in 1988. Currently CERT/CC (http://www.cert.org) is a non-academic and non-governmental unit funded mainly by the U.S. Department of Defence and the Department of Homeland Security along with number of other federal civil agencies and private sector. CERT/CC is a component of the larger CERT Program at the Software Engineering Institute where other areas of work include education and training, research and development, situational awareness, and global relationships. See the CERT FAQ, headings A1-A2, at the web pages of CERT, http://www.cert.org/faq/cert_faq.html [updated 15.12.2005, visited

> many of CSIRT teams and coordinates with them for incident response, but it remains independent of them in terms of organisation. Organisations wishing to use "CERT" in their name must request permission from the CERT.
> With CSIRT the reference is made to a service organisation that is responsible for receiving, reviewing, and responding to computer security incident reports and activity. Their services are usually performed for a defined constituency that could be a parent entity such as a corporation, governmental, or educational organisation, a region or a country, a research network, or a paid client.[592]

## 4.1 Who regulates?

The regulators in the vulnerability disclosure scheme are the external parties reporting the vulnerability information to the vendors and disclosing them to the public[593]. Also coordinators of the vulnerability

---

23.1.2006].

[592] See the Computer Security Incidence Response Team (CSIRT) Frequently Asked Questions (FAQ) point 1 at http://www.cert.org/csirts/csirt_faq.html [last updated 1.2.2002, visited 23.1.2006]. CSIRTs come in a variety of forms. There are, at least, national CSIRTs providing incident handling services to a country or a large region such as the Aus-CERT (The Australian CERT, http://www.auscert.org.au) and CERT-FI (The Finnish CERT, http://www.cert.fi), dedicated internal teams in commercial, governmental or academic organisations providing services to their parent organisation like the DoD-CERT (U.S. Department of Defence CERT, http://www.cert.mil) and the FUNET-CERT (Finnish University and Research Network – Computer Emergency Response Team, http://www.cert.funet.fi), coordination centres facilitating the handling of incidents across several CSIRTs such as the CERT/CC and the US-CERT (U.S. governmental CERT, http://www.us-cert.gov). Many of these teams are members of the Forum of Incident Response and Security Teams (FIRST, http://www.first.org), that has operated as a coalition to exchange information and coordinate response activities among the large variety of specialised groups since it was established 1990. CERT/CC is even a founding member. See the Computer Security Incidence Response Team (CSIRT) Frequently Asked Questions (FAQ) point 4 at http://www.cert.org/csirts/csirt_faq.html [last updated 1.2.2002, visited 23.1.2006].

[593] Of interest are only the benign discoverers. The malign ones may disclose the existence of the vulnerability to the vendor, but typically there is no

reporting process can function as regulators. Both the reporters and the coordination centres act as gatekeepers. They control the message flow and a central informational resource in the development of secure software.[594]

The reporters are a relatively heterogeneous group. A clear source of vulnerability information is found from the organisational customers of the vendors that discover new vulnerabilities during their testing of the fit of the software component for the needs of their information systems. Another central source of vulnerability information is formed by the knowledgeable individual customers that discover vulnerabilities during beta-tests or in other situations where they are probing the software to see how it functions[595]. Also certain hacker groups, in the benign sense, are dedicated to discovering and publishing vulnerabilities, together with other interested parties, such as research institutions, security organisations and interested individual. Even other software vendors that develop compatible software can discover and disclose vulnerabilities.

Central regulators are also the various types of coordinators of the vulnerability reporting process. Certain CERT like institutions even conduct their own vulnerability research and make discoveries by themselves. But their core role is in disclosing the vulnerabilities found by others. They are thus acting mainly as intermediaries and are part of the vertical package of this type of informative regulatory instrument. Thus they are mainly implementers and considered as regulators only for their independent information provision activities. As pure coordinators of vulnerability reporting process they are implementers.

---

intention to improve the security of the software in question. Often they do not publish the knowledge at all. They find vulnerabilities in the purpose of abusing them in order to gain access to systems or to blackmail the vendors.

[594] This is also acknowledged by Tiina Havana in *Communication in the Software Vulnerability Reporting Process*, p. 66.

[595] According to an empirical study over 80% of vulnerability reporters that responded to the survey made by Havana, *Communication in the Software Vulnerability Reporting Process*, p. 31, told that they discovered the vulnerabilities personally in the course of their own work and 60% heard about a vulnerability from an internal testing group.

## 4.2 Influence mechanism

Vulnerability reporting as a regulatory instrument functions by providing information about vulnerabilities. As a regulatory instrument the disclosure of vulnerability information relies on informing vendors, consumers, other users, and citizens more generally, about vulnerabilities in software products and expecting them to act in ways that advance the objective of secure software development. It consists of the communication between different parties under the influence of a specific vulnerability. This type of regulation is not at first hand directed at the primary market activity of supplying goods, but rather at the flow of information necessary for the market to function.

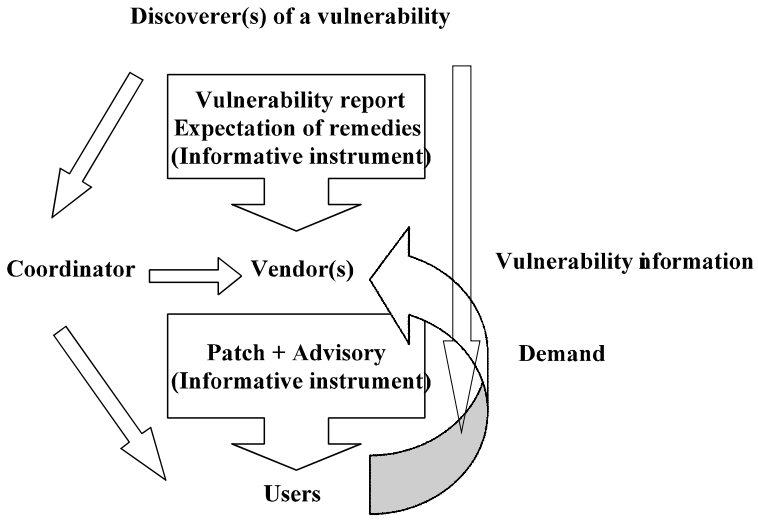
In order to see how this regulatory instruments can influence secure software development it is not enough to state that it, as an informative instruments, covers attempts at influencing people through transfer of knowledge, communication of reasoned argument, and moral suasion in order to achieve a certain widely accepted goal which in this case is the development of secure software. The mechanisms by which the effects are created need to be analysed further.

In relation to secure software development this informative instrument can affect in two basic ways; either through *decisions by vendors* to alter their products in ways that make them less vulnerable due to the increased customer demand or through the *decisions of the users* to alter their behaviour with respect to the risks involved (e.g., to install patches or to take other protective measures). These influence mechanisms find support from the economically minded signalling theory with credence goods[596] and the more general theory

---

[596] For example, using a non-cooperative game-theoretic model with incomplete information Timothy J. Feddersen and Thomas W. Gilligan in Saints and Markets, p. 149-171, show that information-supplying activists can alter the decisions of both firms and consumers on a market for credence goods. Under certain restrictive assumptions, the presence of such informed activists in the market can at least partially mitigate the adverse selection problem and enhance social welfare. They consider the role of activists providing information on a firm's operating practices, such as whether or not it harms the environment, promote employment discrimination, or embrace controversial social or political positions. However, their theoretical

of private politics[597]. Both of them go either directly or indirectly through the information provided to the users and potential buyers of the software. A rough sketch of the influence mechanism is presented in picture 4-1.




Picture 4-1. Vulnerability disclosure as vertical packaging

In this former case, the provision of information to the public, of which the threat of doing so is a part, is assumed to alter the

---

explorations have a wider applicability. Especially the quality aspects of security, i.e., the degree of vulnerability in a software product, bear exactly the same credence characteristics.

[597] David P. Baron in Private Politics, Corporate Social Responsibility, and Integrated Strategy, p. 7-45 develops the influence mechanisms in economic terms from a more general theoretical private politics perspective where interest and activist groups attempt to influence economic activity directly without reliance on public institutions or office holders. He draws especially on the long tradition of research in corporate social responsibility. Note that there is no market failure to correct in the basic model used by Baron and the information is complete even though he also briefly considers situations with incomplete information. However, as Baron notes in 2001 (Private Politics, Corporate Social Responsibility, and Integrated Strategy, p. 42), the study of private politics is in its infancy. The implications for economic activity at the firm, the industry, and the economy levels are largely unexplored. The model Baron presents provides a number of predictions but the need for more research is clear as he points out in a subsequent article presenting the research agenda for private politics in 2003 (Baron, Private Politics, p. 47-53).

behaviour of the vendors indirectly by enabling customers to compare software in terms of its vulnerability. The logic is that when the vendors know that information about the vulnerability of their software (and their competitors' software) is provided (or likely to be provided) to all potential buyers and users, the vendors are motivated to put more secure software on the market and also to tell about it. The increased customer possibilities to compare software products on the basis of their vulnerability serves as an incentive for the vendors to fix the vulnerabilities and to disclose the existing vulnerabilities in order to show responsible behaviour or simply to avoid negative publicity caused by non-action when facing a disclosed vulnerability.

By informing the users and potential buyers of the risks involved in the use of the products in question the potential buyers are enabled to make optimal choices with closer to perfect information. With closer to optimal information the buyers can operate as central controllers of the prices and quality of products by striving rationally after their own private preferences[598]. In relation to software, vulnerability reporting as a regulatory instrument enables potential buyers to assess the quality and security of the software in question and the claims that vendors make about them, and to compare the security of software to a certain degree.

As customers and potential buyers become more aware of the security issues and are better able to articulate their security needs and to assess both the quality and security of the software in question, and the claims that vendors make about them, vendors have to compete to meet those needs. With increased knowledge of the vulnerability of several products the customers can compel the vendor issue patches more quickly and to demand initially secure software

[598] In addition to the enhancement of the market rational behaviour, and at the same time with it, vulnerability reporting can also be seen as protecting the informationally weaker party from entering into highly unfavourable and possibly damaging transactions.

in the future[599]. Less vulnerabilities in released software and quick patching become competitive advantages. Market forces will drive security as a competitive issue due to increased customer demand and software development is expected to start to take security seriously.[600]

Vertical packaging in the sense that mainly software vendors and retailers are used as intermediaries to transmit vulnerability information is at the core of both of these influence mechanisms. In the vulnerability reporting scheme the vendor's information dissemination is enabled by making her aware of the existence of the vulnerability and an incentive to disclose is created by the external discoverer/reporter threatening to disclose directly to the public if the vendor does nothing. In the case of vulnerability reporting the vertical packaging thus happens both by enabling and by requiring the intermediaries to transmit the information further. The disclosure of information by the vendors (acting as second level actors, implementers) is either required or enabled by the regulators[601].

The *enabling* happens simply by disclosing the existence of a vulnerability to the vendor. The provision of information to the vendor serves as an informational resource to which the vendors can

---

[599] Increased customer demand for secure software is enabled by removing at least some of the basic informational problems, i.e., the ability to compare products on the basis of their security.

[600] This obviously requires that alternative products exist in the market, i.e. that buyers have something to compare to and alternatives to choose from. This can as such be problematic due to the competitive structure of the COTS software business characterised with networked effects favouring few solutions and the security related vulnerabilities being an industry wide problem.

[601] In the vertically packed model the capabilities of vendors are harnessed for regulatory purposes. Note that vendors are still not considered as regulators. Instead, they are implementing agents. Even though vendors conduct quality assurance testing as part of the development process and, at least ought to be, the main parties who discover and correct security related vulnerabilities, report them inside their organisation, and also increasingly to the public especially in the form of patches, this information provision as such is not regulation. Their intention is not to achieve certain widely accepted policy goals like more secure software by influencing the behaviour of others. Instead, the intention is to increase the demand and sales of the product.

utilise. Thus, it plays a facilitative role in the testing for vulnerabilities done by the vendors. The vendors are made aware, even with evidence (e.g., test cases or proof of concept code), that vulnerabilities in their software exist, and are thus enabled not just to fix them but also to disclose the information further to their customers.

The *requiring* happens by the external party threatening to disclose the vulnerability directly to the public or more correctly, threatening to do so in the case the vendor does not react to the report given to it. Vulnerability is reported to the vendor and it is accompanied with a threat of public disclosure if the vendor does not react accordingly. This has the practical effect that the vendors have an incentive not just to fix the vulnerability but also to disclose it (e.g., in the form of a patch or a workaround)[602]. The threat does not have to be explicitly made since it is enough that the vendors are aware that the public disclosure is a possibility. In addition to case by case expression, such an expectation is also made by the procedures followed by the stakeholders in the vulnerability reporting process[603].

The threat of public disclosure adds the additional non-legal consequence of negative publicity to the inaction of the vendor in the case of vulnerabilities existing in its software. The negative information about products insecurity can damage the business reputation of the vendor. If this negative information spreads among

---

[602] Note that no specific info about the vulnerability or even its existence might be disclosed when releasing a patch or a workaround. However, this typically is done in order to show the importance of the fix to the affected parties.

[603] Due to the conflicting objectives the substance of the policies, the rules and procedures set, differ according to the party whose interest are represented. However, rules on conflict resolution between the vendor and the finder are typically included in the policies for vulnerability reporting. These rules differ, but typically stipulate or at least imply the possibility to exit the process of following the policy or the guideline and thus to publish the vulnerability or to shut down communication. For example the OIS (Organization for Internet Safety) *Guidelines for Security Vulnerability Reporting and Response* include rules on exiting the process in the situation of irreconcilable disagreements. However requirements of reasonable efforts to solve the problem and noticing other parties before exiting are provided together with the possible use a third party to solve the problems.

market actors, they can be discouraged from purchasing the insecure product. Basically, the potential disclosure of vulnerabilities by external parties serves the function of disseminating information about the trustworthiness of the software product and facilitates the application of non-legal sanctions like refusing to purchase[604]. The threat of publication of the existence of a vulnerability in a specific software attaches a non-legal sanction, with consequent loss of confidence and trust, to the report made to the vendor by the external party. In the COTS software markets this can be an influential mechanism since there are not a lot of other sources of trustworthiness of the vendor and the security of its software beside business reputation.

This is actually just another way of employing (implementing, enforcing, carrying out) the public disclosure as an informative regulatory instrument; it is not a separate type of instrument. However, it is the threat that creates the vendors incentive to disclose. Even though the direct disclosure to the public, without notifying the vendor first and thus threatening her with the possibility of public disclosure, can also provide incentives for the vendor both to patch and to disclose the vulnerability together with the patch to the public due to the possible customer outcry and negative publicity if the

---

[604] Hugh Collins in *Regulating Contracts*, p. 124, notes, when discussing the role of the courts of law in collecting and disseminating what participants in the market regard as reliable information about trustworthiness of other contracting parties, that "[t]his function of the legal process need not be performed by an institution which possesses all the trappings and powers of a modern court. What is required is rather a body that is respected by the trading community as an impartial and reliable finder of fact or truth. … In particular, the body does not require the power to impose effective sanctions itself. Its influence upon contractual behaviour stems instead from the independent reactions of participants in the market to its authoritative judgements about business reputation". Even though the external reporters in many cases cannot be seen as impartial and reliable finders of fact or truth or as authoritative parties to judge the issue, the fact that they get their message heard is enough to influence in a similar, even though necessarily in a weaker, manner. In highly competitive markets only a hint of such a sensitive subject as the possible vulnerability of software might be enough to evoke an adverse reaction from customers and potential buyers.

vendor is unresponsive[605], the consensus seems to be that the vendor is first given the opportunity to issue a patch. Under this consensus based responsible disclosure scheme, the incentive for the vendor is created by the threat of public disclosure. The disclosure is a threat, not a weapon. The reporting to the public by the external party after the vendor has been unresponsive further strengthens this effect, but it does not create it[606].

Despite the differing interests that are discussed below and the still somewhat varying approaches to vulnerability reporting, a consensus seems to exist on giving the vendor first enough time to develop a patch; the publication of some part of information (no consensus on the amount) after a predefined time is seen to be the ethically correct way to handle vulnerabilities[607]. The delaying of the

---

[605] Compare to the logic of full and immediate disclosure. The full disclosure movement is based on the assumption that when the disclosure is made directly to the public (even before, or at the same time as to the vendors), the vendors, in the fear of wider negative publicity if it does nothing, is compelled to issue patches quickly. The rationale is that embarrassment pressures software vendors into producing more secure software. There is also preliminary empirical evidence that supports the assumption that vendors are more likely to patch under instantaneous public disclosure and do it faster. See Arora et al., *Impact of Vulnerability Disclosure and Patch Availability*, p. 2. However, this empirical data does not directly speak in favour of such an embarrassment effect. It simply shows the difference in the likelihood and speed of patching.

[606] This is similar to the liability based duty to disclose where the threat of liability has the practical effect that the vendors have a motivation to disclose vulnerabilities in order to shield themselves from liability. See heading 5.2 below.

[607] This is suggested by Havana and Röning in Communication in the Software Vulnerability Process, p. 7. This is also visible in the guidelines trying to establish a common approach, of which the most prevalent, even though their wide acceptance by the security community remains to be seen, are the OIS (Organization for Internet Safety) *Guidelines for Security Vulnerability Reporting and Response* and the NIAC (National Infrastructure Advisory Council of the U.S.) *Vulnerability Disclosure Framework* as presented by Chambers and Thompson. However, there is no consensus on the time that should be given to the vendor. The guidelines leave it to be agreed on a case by case bases between the stakeholders (typically the finder and the vendor; a coordinator is taken abroad when the problems occur). Certain stakeholders have adopted more formal approaches. For example, CERT/CC typically notifies vendors first

reporting to the wider public is considered necessary in order to minimise the risk posed by also the increased possibilities for exploitation before a patch correcting the vulnerability is available.

The consensus approach to vulnerability reporting favour the publication of some information after a predetermined time (responsible disclosure) in order to minimise the negative effects of increased vulnerability of the customers and the increase in the number of attacks. The immediate, full disclosure like, effect on the user ability to protect oneself is postponed in order to minimise the risk of exposure to attack due to the existence of the vulnerability spreading before a patch is available. This is seen as a balance between the dual natures of vulnerability disclosure both in helping protection and easing attacks. At the same time it is a central limitation to the scientist's freedom to decide about the publication of research results, which is a part of the freedom of science guaranteed in the constitutions of many countries[608], in the name of the security of the information infrastructure. This consensus approach also seems to be more optimal in societal terms[609].

Vertical packaging, the use of vendors as intermediaries in the information transmission, is only a part of the influence mechanism of vulnerability disclosure concerning secure software development. In addition to the use of intermediaries, vulnerability disclosure also

---

after getting the vulnerability reports and provides vendors 45 days to patch the vulnerability before making the vulnerabilities public. Further, in almost all cases, CERT/CC discloses information about vulnerability only after vendors issue patches fixing the vulnerability. CERT/CC does not release all technical information about a vulnerability to the public.

[608] See, for example, section 16 subsection 3 of The Constitution of Finland (731/1999). For a discussion and a review of the constitutional doctrines of other countries, see Miettinen, *Tieteen vapaus*.

[609] This is what Ashish A, Telang R and Xu H (2004) *Timing Disclosure of Software Vulnerability for Optimal Social Welfare*, Carnegie Mellon University working paper, April 2004, suggest by using a game theoretic model (referred to in Arora et al., *Impact of Vulnerability Disclosure and Patch Availability*, p. 2).

involves direct supply of information to the users and to the public at large.

In this latter form of influence mechanism, the information provided to users of the existence of a vulnerability in a software product, its severity, etc. enables them to install patches or take other protective measures to defend themselves when needed. The reporting of vulnerabilities to the public is considered to encourage people to make sure their software is up-to-date. It is assumed that when the users are provided with the knowledge of the vulnerability they are at least enabled to take protective measures. To the degree they do so, the damage caused by the abuse of the defective software products is diminished. This is clearly visible in the rhetoric's of those favouring the reporting of vulnerabilities by external parties. The purpose in vulnerability reporting is considered to be to provide to the users of the software what the attackers already are assumed to know (i.e., the existence of the vulnerability), and thus to enable them to defend themselves.

The direct provision of information to the public by the reporters, together with the threat of doing so, not only attempts to alter the behaviour of the vendors indirectly by enabling customers to compare software in terms of its vulnerability. It also attempts to alter the behaviour of the users and potential buyers of the software directly without intermediaries. The users are enabled to alter their behaviour with respect to the risks involved even without vendor influence. This direct provision of information can circumvent the vendor disincentives for disclosure and put the information available quickly when needed.

Even though the external constraining depicted above is the main type of influence, intrinsic predisposition shaping is also included. In addition to the effort to alter the preferences of (potential) users towards more security consciousness in the purchasing of COTS software by informing them about vulnerabilities, similar efforts to alter the intrinsic predispositions both of the vendors and their customers are made when the public disclosure of vulnerability is presented as an ideological argument of the freedom of information

and its positive relationship to security. This is most clearly visible in the immediate public disclosure form, but is in the background of the whole vulnerability disclosure scheme.

## 4.3 Factors shaping the influence

### 4.3.1 Objectives

As a regulatory mechanism, vulnerability disclosure serves the objective of influencing secure software development. The disclosure of vulnerabilities arose from the need to inform the vendors about vulnerabilities in their software products. The need to inform the public at large and to threaten the vendors with of disclosure of vulnerabilities to the wider public has been considered to become issues due to the unresponsiveness of the vendors. As such, the improvement of the security of existing software is an explicit regulatory objective.

Information on the objectives of the main regulators, the reporters of vulnerabilities, is scarce. Preliminary research shows that the vendors and the reporters tend to explicitly state, as they did in answering to an questionnaire about the values and beliefs that guide their decisions related to vulnerabilities, that security is the most important value in the reporting process[610]. In the end, however, their reasons for the interest in security are different. Whereas the reporters seem to strive for a level of security that is beneficial for the general public[611], the vendors are interested in the level of security that fulfils only their customers' needs. This is what they are expected to do under normal business logic. As the study of Tiina Havana and Juha Röning suggest, the public benefit and the public's right to know are valued more by the reporters than the vendors[612].

---

[610] Havana, *Communication in the Software Vulnerability Reporting Process*, p. 70.

[611] To the reporters the communication process, and the reporting as such, is for the benefit of the society.

[612] Havana and Röning, Communication in the Software Vulnerability Process, p. 7 and 11. The reporters also value more open information transmission.

Still their objectives are not fully unselfish or purely altruistic. In addition, the objective of the reporters might not be favourable to the secure software development process. For example, they can lead to disclosures that unnecessarily increase the risk caused by unfixed vulnerabilities.

An objective of the reporters, according to anecdotal evidence especially from the receiving organisation, is to gain reputation that can be cashed later on as increased job opportunities. Recognition and fame are relatively easily achieved due to the existing high media interest and the admiration from an adoring and grateful public of the unselfish geniuses that are seen to be spending their valuable time in benefiting the public. The reporting might be used as marketing to highlight the technical skills of the reporter. Thus, there is a possibility that the reporters might be less interested in the results of the disclosure to the public than in the visibility of their actions[613]. Even thought this is pure speculation, as a possibility it cannot be disregarded. Though, without further evidence it cannot be given much significance.

The feasibility of this informative instrument can partly be explained by the final influencing objective of information on secure software development being largely hidden. It is politically easier to give people better information about risks than to get companies to stop putting people at risk directly. At the same time the regulators commitment to educated choice and informed citizens is elicited.[614]

---

This conclusion is based on higher support for full and immediate reporting among the reporters than the vendors reported ibid, even though the consensus seems to be on partial disclosure after a predefined time.

[613] This is in line with the argument made in political studies concerning information as a tool of government that due to the relative cheapness of the use of informative instruments and their visibility politicians may be less interested in results than they are in showing that they are doing something. See, e.g., Weiss, Public Information, p. 234.

[614] Of these political reasons especially for using public information as a tool of government, see Weiss, Public Information, p. 235.

### 4.3.2 Substance[615]

The substance of vulnerability reporting is the provision of knowledge of a vulnerability to the various affected parties. Vulnerabilities are reported both to the persons or organisations responsible for fixing or distributing the knowledge further to other relevant parties (i.e., to vendors or to mediators like the CERT), and to the public directly.

The consensus among stakeholders in the vulnerability reporting process is that the report should initially be made to the vendor and that the vendor should first be given enough time to develop a patch. If the vendor fails to issue a patch or otherwise fix the vulnerability after a sufficient timeframe, which may vary depending on the type of vulnerability and be agreed upon case by case basis between the reporter, receiver, and possible the coordinator, or be fixed 30-45 days, the reporter may go public with the vulnerability thus giving users a possibility to defend themselves. As already discussed, the publication of some part of information (no consensus on the amount) after a predefined time is seen to be the ethically correct way to handle vulnerabilities.

The disclosure to the public is either done by the vendor in a way that shows her responsible behaviour, by the reporter directly, or both[616]. The difference is that when the vendor is involved, her responsible behaviour is a positive sign for the customers and further negative publicity due to the vulnerability is diminished. The purpose

---

[615] The substance and the implementation of vulnerability disclosure are difficult to separate. This is due to the reporters operating at both stages. Most of the issues are considered as implementation and analysed below. In here, only tentative points that relate to the margin of discretion of the reporters are raised.

[616] The threat of public disclosure that completes the disclosure requirement is thus typically time limited; in the ethically correct way of first reporting to the vendor, only a grace period is given, after which the disclosure to the public is made even though the vendor already might have issued a patch. However, in this way the vendors responsible behaviour is shown and negative publicity downsized. By disclosing the vulnerability to the public in any case the awareness of the existence of the vulnerability and of the possible patch is increased and a way for the reporter to gain reputational credit is provided.

is to get the knowledge of the existence of the vulnerability, and a fix to it, to the affected parties so that they can take protective measures. If the vendors do not do that voluntarily after they have been notified of the existence of the vulnerability, then the reporters do it any way[617].

The coverage of vulnerability reporting is central to its effectiveness. When the concern is only on disclosing vulnerabilities found after the release of a software package to the public, the inherent limitation of the substance of vulnerability disclosure becomes apparent. It affects secure development only in terms of patching (i.e., post the release of the software). Initial secure software development is not the primary concern. What this means is that vulnerability disclosure is just a supportive mechanism for initial secure software development; it does not replace it and should not be used as an excuse to do so. This is so even though in the current state of immaturity in secure software development it has gained much attention and a central role as a regulatory instrument. This limitation of the coverage of vulnerability disclosure needs to be explicitly stated even though it is obvious for those participating in the reporting process[618].

There are, however, at least three ways in which vulnerability reporting can influence initial secure software development in addition

---

[617] The direct disclosure to the public by the external discoverers of the vulnerability can happen even if the affected vendor has fixed the bug and disclosed its existence to its customers. The emphasis on the reporters direct information provision is visible especially in the procedural rules where the reporters are given the possibility to publish vulnerabilities that are patched but not made public by the vendor. Such a guideline for unannounced fixes is given, as presented by Chambers and Thompson in *Vulnerability Disclosure Framework*, p. 31, at least in the NIAC guidelines for the discoverers (reporters) of the vulnerability.

[618] The limitation has been pointed out, e.g., by Quirchmayr et al. in A Business Process Engineering Based Approach towards Incorporating Security in the Design of Global Information Systems, p. 1115. It should be noted that the proponents of vulnerability disclosure does not, as a default, expect the reporting be a cure for all. Instead, it is, as it should be, considered only as a supportive instrument. Not even the vendors that increasingly rely on vulnerability disclosure seem to expect it to be a cure for all.

to post-release correction of software[619]. In the first, historical data about the vulnerability of a specific software package and about the responses its vendor has given to the reporters can be useful in software acquisition. At least it can be useful in considering the switch to competing products. To the degree the responsiveness of the vendor to the reports become public, as they do when the disclosure is made directly to the public if the vendor is not responding, information about the trustworthiness of the vendor is also disseminated. Even though this data does not fully reflect the security of next release of the software package in question, it can be used as indication of the seller's attitude towards secure software development in general.

In the second, vulnerability reporting can influence initial secure software development more forcefully if the information provided is internalised by the vendors. This would require the vendors to distribute the information about discovered defects to their software developers and to make that information an essential part in the software development process. Currently this is not commonly done[620].

In the third form, vulnerability reporting can influence initial secure software development also when knowledge of a vulnerability, not just of its existence but also of its details (like the specific coding or implementation error that causes it) is used to train current, and educate future, software developers more widely to avoid the causes

---

[619] Whether or not it does so depends heavily on the ways the users and the vendors utilise the provided information. This is why these issues are discussed further in the implementation and reaction of objects sections. Only a short sketch is provided here.

[620] Only a little over half of the receivers of the reports answered in a survey conducted by Havana, as reported in Havana and Röning, Communication in the Software Vulnerability Process, p. 8, that they pass the information further to their software developers with the intention of preventing similar vulnerabilities occurring again. Of all the respondents to the survey only 15 per cent do this, and when they do it, the information does not have an essential role in the software development process. Thus, the information gathered during the reporting process is not widely utilised for the benefit of secure software development.

of the vulnerabilities. Due to the current lack of teaching material such knowledge, especially if codified, could serve as an important educational resource. For example, recent textbooks on secure software development use known vulnerabilities especially as examples to clarify and make things more concrete[621].

Another issue in the substance of vulnerability reporting is the use of public disclosure as a punishment for vendor misbehaviour. As already shown, the threatening with public disclosure is a central part of the effect mechanism of vulnerability reporting. Even though this threat of punishment in the form of negative publicity essentially makes the reporting effective in the final end by forcing the non-cooperative vendors to comply with requirements of the reporter about fixing the bug and disclosing its existence to the customers, it can also have adverse effects.

When the substance is partly based on the assumption that the vendors cannot be trusted, that they have to be threatened with sanctions in order to correct vulnerabilities in their software, instead of voluntary compliance, the juxtaposition of vendors and the reporters with the general public is emphasised. This might hamper cooperation especially between the vendors and the reporters. To a certain degree it has already done so as the heated ideological discussion of full and immediate disclosure described. The basis for cooperative approaches took a while to be found. It still influences the cooperation as the difficulties in finding a common cooperative approach to the reporting process that all stakeholders could approve show.

However, the use of coercive power by the reporter in the form using public disclosure as a punishment for vendor misbehaviour is not the only influence mechanism external reporters' use. Neither is it the mostly used, at least not anymore when full and immediate disclosure has faded into margin. The use of less coercive and more cooperative influencing mechanism like mere informing of the

---

[621] See, e.g., Graff and van Wyk, *Secure Coding*, from 2003 and Viega and McGraw, *Building Secure Software*, from 2002.

existence of the vulnerability and persuasion are used. The public disclosure is only the last resort when cooperation and persuasion no longer work.

A central question in the substance of vulnerability reporting is the wide discretionary power of the reporters, of which the various approaches to vulnerability disclosure even among the reporters is an evidence (e.g., from full and immediate disclosure to non-disclosure). As vulnerability reporting has matured and become more clearly an established part of software development process, this freedom has turned out to be a hindrance for the effectiveness of reporting and for the codification and unification of the practices in the reporting process[622].

The policies and guidelines for vulnerability reporting that codify and unify the approaches and thus limit the freedom of the reporters, the influence of which are analysed below in the section concerning implementation, still leave considerable margin of discretion for the reporters. The policies and guidelines set forth by various stakeholders are not coercive even though the receivers of the reports are likely to be more cooperative and even more responsive when their reporting policies are followed.

What information is disclosed by the discoverer of a vulnerability and the amount of it influences the effects vulnerability reporting can have. While reporting to the vendor (instead of directly to the public), possibly even via a coordinator, the amount of details affects first of all vendor behaviour. More details not only increase the credibility of the report but also ease the verification its accuracy. This makes the vendors more likely to respond accordingly. At least all technical information together with information of the discoverers environment (e.g., hardware, configuration, other applications installed, relevant details about the network topology, firewall rules) and the level of

[622] The necessary communication between the stakeholders of vulnerability disclosure requires codified rules which all the parties can accept.

patching are considered relevant for the vendor to be able to verify the accuracy of the reports[623].

In order to prevent premature leaks of vulnerability information, that imposes increased risks to users, the protection of the sensitive information is important. However, with restricted reporter-vendor relationship this can relatively easily be done[624].

The amount details have much more varying effects when the reporting goes directly from the discoverers to the users. The level of detailed information made available to the wider public can have conflicting consequences due to the information being useful both in correcting and exploiting vulnerabilities. The details not only give precise information for the vendor to verify the claim and to create a patch, but also for the user of the software to verify that the vulnerability exists in her systems and what actions could be taken if the vendor has not offered a patch. However, if the details become more widely known, e.g., in the case of disclosure via full disclosure list like the Bugtrack[625], also the attacker is helped at the same time in the creation of exploits[626], in the determination of vulnerable systems, and in the launching of attacks.

This is why, the disclosure of all details is not considered appropriate even in situations where the vendor does not inform its users of a fixed vulnerability or is otherwise unresponsive to the

---

[623] Chambers and Thompson, *Vulnerability Disclosure Framework*, p. 29.

[624] If, however, premature leaks occur, or the vulnerability otherwise becomes under active exploitation, the guidelines typically recommend certain amount of coordination between the discoverer and the vendor. At least in the form of informing the other party and providing evidence that the information has already been made public or that active exploitation is under way in order not to worsen the situation. (OIS, *Guidelines for Security Vulnerability Reporting and Response*, p. 19, point 8.4; Chambers and Thompson, *Vulnerability Disclosure Framework*, p. 30).

[625] The direct reporting to the users happens mostly by mailing lists like the Bugtrack or even popular media.

[626] As Chambers and Thompson point out in *Vulnerability Disclosure Framework*, p. 30, with full details it is relatively easy for fairly unskilled malicious individuals to develop exploits thus increasing the immediate risk to end users.

reports, and the formal vulnerability reporting process is aborted. The consensus in reporting the vulnerability information directly to the public as a last resort (when all attempts to work with a vendor directly and through a coordinator fail) seems to be that full details are not disclosed. Instead, a balance is considered more appropriate[627].

Without going deeper in to the effects of explicit details (e.g., how the publication of the proof of concept code at source level influences) it can be stated on the basis of current research that automation of a vulnerability, not just its disclosure, serves as the catalyst for widespread intrusions[628]. While the explicit details serve both the correcting of vulnerabilities, their exploitations and the defence of information systems, the automation of vulnerability exploits seems to serve only the exploitation of vulnerabilities.

Another issue in the substance of vulnerability reporting, which influences the effects vulnerability disclosure can have, is the basic choice in the use of different reporting channels; either directly to the vendor (e.g., its security experts or product support), or through a coordinator like the different CSIRT organisations or the CERTs[629]. This is typically left to the discretion of the reporter in the policies and guidelines. Only the requirement that the vendor is first given sufficient time to issue a patch is seen as a common approach; immediate and full disclosure is seen as unethical.

The different levels of credibility these channels possess in the eyes of the vendors receiving the reports affects the effectiveness of reporting[630]. Software vendors with tight schedules quickly need to

---

[627] Chambers and Thompson, *Vulnerability Disclosure Framework*, p. 30. In the current consensus based policies the achievement of this balance is left to the discoverer.

[628] See Arbaugh et al., Windows of Vulnerability, p. 57.

[629] For definitions and background, see above p. 162-163.

[630] As Paul Slovic, one of the leading analyst of risk, risk perception and risk management, emphasise in Perceived Risk, Trust and Democracy, p. 318-319, while discussing the importance of trust in risk management, trust towards the manager of risk and into the source of risk information is fundamental for the

verify the existence and seriousness of the reported vulnerabilities and thus seek for sources of vulnerability information that are both reliable and credible. This is especially what established independent third-party CERT like organisations offer.

The reporters, instead, easily lack credibility in the eyes of the vendor as a source of vulnerability information. Reports are often perceived, especially by the vendors, as the private advertising of the competence of the discoverer. Perceiving the competence advertising and reputation gaining as primary reasons for the reporting might affect the effectiveness of vulnerability disclosure seriously. Not only are the vendors less likely to consider the reported information credible, but also is their overall interest in the vulnerability disclosure in general likely to decline. The assumption often made that the reporters act on the basis of dislike of the vendor (e.g., it being the "monopolist", a major international operator) does not ease the issue[631].

How common this kind of reporting in the private interest of the discoverer of the vulnerability is, is an empirical question of which no certainty currently exists. The objective of personal reputational gains of the reporters overriding everything else is often raised especially in the cases where a direct public disclosure is made without any advance verification of its accuracy by the vendor (i.e., full and immediate disclosure)[632]. However, there are important theoretical

---

effectiveness of risk-communication. According to Slovic, idem, p. 318, the limited effectiveness of risk-communication efforts can be attributed to the lack of trust.

[631] Research has identified certain mismatch in the channels used by the reporters and the information sources utilised by the vendors. Whereas the former seem to be relying on direct contact with the vendor (e.g., via email, bug reporting forms), the latter utilise coordinators as their primary source of vulnerability information instead of the direct reports. This is visible in the first empirical study of the communication processes in vulnerability reporting conducted by Tiina Havana, *Communication in the Software Vulnerability Reporting Process*, p. 42. One possible explanation for this mismatch is the credibility of the information source in the eyes of the vendor.

[632] See, e.g., Chambers and Thompson, *Vulnerability Disclosure Framework*, p. 23-24.

assumptions that speak in disfavour of the likelihood of inaccurate reports.

Since the reputation of a reporter, the primary asset asserted to be sought by the reporters, is likely to be rather quickly loosed if the disclosure is found to be groundless in later analysis, short term reputational gain-seeking is not likely to lead to high amounts of inaccurate reports. The huge amount of disclosed vulnerabilities further diminishes both the possibilities to gain reputation and to get caught of inaccurate reports. And if the reporters take seriously the negative impact of large numbers of inaccurate reports and act responsibly, the likelihood of inaccurate reports is further diminished. The assumption thus would favour the reliability of reporters as sources of vulnerability information. This is so despite the wide unavailability of traditional legal disincentives for false or misleading statements due to the anonymity of many of the reporters, whish only increases with the use of unauthorised versions of the tested software, and the reporters being located in different jurisdictions.

Still, the use of mediators in the reporting process can influence the likelihood that the reports have the desired effects on the behaviour of the vendors (i.e., they are taken seriously and corrective action is taken). Reporting through an established mediator not only can increase the credibility of the report in the eyes the vendor, and thus become more likely to be taken seriously, but it also increases the likelihood of the report to reach a wider number of the affected users. Especially in the case of established and networked national and international CERT or CSIRT organisations. The implementation of vulnerability reporting as a regulatory measure further depends on the channel used since the practices and procedures followed in reporting differ between the actors. This will be discussed further in the analysis of the implementation of vulnerability reporting.

### 4.3.3 Implementation

As noted, the disclosure rules are vertical packages where mainly the threat of external public disclosure is used to incite and obligate the vendors to disclose vulnerability information (upper level of package), and then the vendors inform the users (lower level of package). Since the implementation in the upper level of package is conducted by separate institutions and focuses on different types of instruments (written and informative), whereas in the lower level of package the implementation is conducted by the vendor and concerns similar informative instruments, it is useful to analyse the effects of implementation at both levels separately. The upper vertical level is analysed first.

*Upper vertical level.* Publishing vulnerability information to the public, enforcing the threat of public disclosure (the core of the effect mechanism of vulnerability disclosure) is a central part of the implementation of vulnerability reporting as a regulatory instrument. Deterrence is a central, even though not the only part of the effect mechanism. The actual disclosure of vulnerability information to the public in cases where the vendor is not responsive is essential for the effectiveness of the deterrence mechanism[633].

Even though both the vendors and the reporters tend to explicitly state, as they did in answering to an questionnaire about the values and beliefs that guide their decisions related to vulnerabilities, that security is the most important value in the reporting process[634], the role of the deterrent effect has been important. This is because, in the end, their reasons for the interest in security are different. Whereas the reporters seem to strive for a level of security that is beneficial for the general public[635], the vendors are interested in the level of security that fulfils only their customers' needs. This is what they are

---

[633] This possibility is left open also in the consensus seeking guidelines as has been noted above.

[634] Havana, *Communication in the Software Vulnerability Reporting Process*, p. 70.

[635] To the reporters the communication process, and the reporting as such, is for the benefit of the society.

expected to do under normal business logic. As the study of Tiina Havana and Juha Röning suggest, the public benefit and the public's right to know are valued more by the reporters than the vendors[636].

Instead of having the objective to gain a level of security that is sufficient for the society as a whole, the vendors' objective with reporting policies seems to be formalisation and cost-effective management combined with the private intention to increase the predictability and control over the process. At best, the intention is to achieve cost-effective help in security development from the companies' point of view and the preservation of the public image of the company and its products. The vendors seek to fulfil the expectations of their stakeholders about the products and the company.[637]

> Because users are increasingly enlisted in finding vulnerabilities[638], vendors have a need to manage the information that may damage

---

[636] Havana and Röning, Communication in the Software Vulnerability Process, p. 7 and 11.

[637] This is also suggested by Tiina Havana in *Communication in the Software Vulnerability Reporting Process*, p. 70. This difference in objectives may partially explain the difference in the use of formal policies pointed out by empirical research. Whereas nearly half of the receiving organizations had some kind of a reporting policy during an empirical study in 2002 reported by Havana in *Communication in the Software Vulnerability Reporting Process*, p. 72, only one third of the reporting organizations had one. Standardised procedures in terms of policies and guidelines are more important to the vendors due to the high economic interests involved.

[638] Vendors are increasingly turning to their customers for help in finding defects in their software by making early releases (pre-release or beta versions) available to interested users and by distributing incremental updates (e.g., patches) often to the software. This applies especially to the open source software vendors, but also the proprietary software developers understand just as well the value and potential of users in testing. The principle of release early and often makes economic sense also for the proprietary vendor. Proprietary vendors desire to achieve more cost-effective testing by externalising part of the costs of testing to the customers. The economic rationale for the increased reliance on vulnerability management and patching is analysed, under the assumption of the availability of full quality information for the customers, by Ashish et al. in *Sell First Fix Later*.

the public image of the company, and at the same time make users even more vulnerable. The vulnerability management and reporting policies help in this by minimising the negative side effects (both in terms of increased vulnerability of users and especially of negative public image) and by speeding up the vulnerability management, and thus decreasing the costs involved.[639]The vendors see the communication in the software vulnerability reporting process as a management tool and a resource for the development of secure software; from their perspective the information transmission should be codified and organised as pointed out by Tiina Havana and Juha Röning[640]. Their objective in setting up the vulnerability reporting policies is partly to improve their products by managing the flow of information by codifying and making explicit the procedural knowledge and by enhancing the dialogical communication between the vendors and the reporters[641].

Both the reporters and the vendors objective is to avoid harming others (to minimise the users window of vulnerability and to avoid the widespread abuses of the vulnerabilities that often follow a premature disclosure), but the vendors are more concerned with avoiding fear, uncertainty and doubt[642]. Even though this is in line with the societal interest of enhancing e-commerce, and naturally with the private interest of the companies, this may also be interpreted as the vendor's objective to avoid negative publicity by restricting the amount of information published and the time of publication by setting up policies for vulnerability reporting[643]. This objective

---

[639] The economic rationale for the increased reliance on vulnerability management and patching is analysed, under the assumption of the availability of full quality information for the customers, by Ashish et al. in *Sell First Fix Later*.

[640] Havana and Röning, Communication in the Software Vulnerability Process, p. 10.

[641] Havana, *Communication in the Software Vulnerability Reporting Process*, p. 68-69. More intensive dialog between the reporters and the vendors, instead of mere one-way communication, has been deemed necessary to increase the validity of the reports as noted by Tiina Havana in *Communication in the Software Vulnerability Reporting Process*, p. 69.

[642] Havana and Röning, Communication in the Software Vulnerability Process, p. 7.

[643] This finds support from the empirical results reported by Havana and Röning in Communication in the Software Vulnerability Process, p. 9, that the

> combines with the objective of avoiding possible legal actions on the basis of liability of the vulnerabilities. Also the objective of gaining favourable publicity by appearing to be a responsible actor can have an effect, since the adoption of vulnerability reporting policies shows that they are at least doing something and taking the reports seriously.

As a deterrent, however, the (threat of) public disclosure of vulnerabilities faces serious hindrances. As empirical research on deterrence suggest, it is certainty of punishment that is more important for deterring crime rates for more 'rational' crimes (as opposed to spontaneous emotional crimes) than the severity of punishment[644]. Severity of punishment only has a deterrent effect when the certainty level is high enough to make severity salient. In the case of threatening with public disclosure of vulnerabilities the severity of the non-legal sanctions might be enough as such to create a deterrent effect, but due to randomness of the public disclosure of vulnerabilities, it necessarily does not provide enough deterrence for software vendors to take security seriously. However, with the increasing interest in vulnerabilities the low certainty of punishment in the form of public disclosure might be diminishing.

This form of enforcement through deterrence can influence the possibilities for cooperation between vendors and reporters, which is crucial for the effectiveness of vulnerability reporting. When the implementation is partly based on the assumption that the vendors cannot be trusted, that they have to be threatened with sanctions in order to correct vulnerabilities in their software, instead of voluntary compliance, the juxtaposition of vendors and the reporters with the general public is emphasised. This might hamper cooperation

---

vendors oppose full disclosure more than the reporters even when it is done after a predetermined time, even though full and immediate disclosure is opposed by both parties (vendors and reporters). Perhaps the condescending attitude towards vulnerabilities is an attempt to hide the intentionally increased reliance on patching and on other post-release quality improvement activities.

[644] This is pointed out by Antunes and Hunt in The Impact of Certainty and Severity of Punishment, p. 189-190, in their summary of conclusions from empirical research on deterrence in the U.S. before 1980.

especially of those vendors that voluntarily would correct the vulnerabilities, issue patches and inform their customers[645]. To a certain degree it might have already done so, as the heated ideological discussion of full and immediate disclosure described. The room for cooperative approaches was lost for a while.

Currently the implementation of vulnerability disclosure has turned from sanction based full disclosure towards forms of responsible disclosure where negotiated solutions and cooperative correction of security related bugs are primary. This is what the agreed consensus-based formal processes codified into the policies and guidelines are signs of. The public disclosure of a vulnerability is currently a result of cooperative remedying by the vendor, reporter and, when needed, the coordinator. The threat of public disclosure if the vendor does not issue a patch is enforced only as a last resort; when all attempts to work with a vendor directly and through a coordinator fail[646].

This cooperative mode of implementation is likely to be efficacious in the case of vulnerability disclosure because the information reporters provide to vendors bring mutual benefits[647]. The vendors get low cost assistance in testing security and the reporters get more secure software together with fame and possible future benefits. The cooperative implementation also appeals to the motives of vendors

---

[645] Of this weakness of the deterrence mode of enforcement in general, see Ayres and Braithwaite, *Responsive Regulation*, p. 25.

[646] This is visible especially in the consensus based approaches to vulnerability reporting, such as the NIAC guidelines. See Chambers and Thompson, *Vulnerability Disclosure Framework*, p. 30. This is inline with the argument in the theory of private politics emphasised by David P. Baron in Private Politics, p. 36, that many firms attempt to avoid private political actions by proactively adopting policies that reduce the likelihood that they become a target. Even though some of these attempts are little more than public relations, many of them represent real commitments to changes in policies and practices. This is also why the number of public disclosures or other observed conflicts do not tell the truth about the influence of vulnerability disclosure.

[647] The requirements for effective cooperative regulatory enforcement are informally adapted from those presented in the context of agency based implementation. See, e.g., Scholz, Cooperative Regulatory Enforcement and the Politics of Administrative Effectiveness, p. 115-136.

to act socially responsibly in the case security related vulnerabilities occur. They get a chance to show to be acting responsibly by reacting quickly to the reports without necessarily actually changing the initial secure development practices.

Vendor unresponsiveness to vulnerability reports and reluctance to take them seriously, which was one of the arguments used to justify immediate public disclosure, is still a serious problem for this cooperative implementation strategy; even so when mutual benefits can be gained. For example, a fear of legal actions under U.S Digital Millennium Copyright Act, EU Copyright Directive or similar national laws has led independent security researchers (i.e., those not in direct cont(r)act with the vendor in question) to withdraw from reporting[648].

The consequences of this possible legal threat to the security research and vulnerability disclosure are real despite of the suspicion of the validity of the liability claims under many circumstances. Just that the legal rules are written in a manner that makes them really hard to understand and open to interpretations provide sufficient basis for this fear. It is no wonder that security researchers are cautious, when the rules have already been used to threaten certain researchers.

The problem is that only long-term benefits are involved and require longer term objectives than currently exists in the market driven software development practice characteristic of COTS software. In the short run, both the vendors and the reporters would be better of with not cooperating. The reporter would get, e.g., immediate and wide media attention due to the public disclosure[649] and the vendors would not have to consume resources like labour and time to fixing

---

[648] Many security reporters, like Ed Felten, have reported having withdrawn from reporting a vulnerability to the public due to the legal risks and threats of legal action by the vendors. A general concern has also been reported in 2004 by John T. Chambers and John W. Thompson in the final report for the U.S. National Infrastructure Advisory Council (NIAC), *Vulnerability Disclosure Framework*, p. 25.

[649] Even though this is questionable in the current situation where responsible behaviour is expected from all the parties and initial contact with the vendor is considered appropriate, the reporters still would get more chances to show their capability; at least outside the professional community.

the vulnerabilities and could concentrate on getting the next version to market quickly.

A long-term relationship between the reporters and vendors mitigate this problem[650]. When cooperation around a vulnerability between reporters and vendors possibly together with other vendors and coordinators is repetitive, parties become sufficiently concerned with future encounters to resist the short-term temptation to shirk. Even though the encounter might not be with exactly the same actors, the publicity of shirking in the current high media interest atmosphere alone is sufficient to transmit the information to other reporters and vendors[651].

Due to its dual nature as an informative instrument, i.e., there is an element of the direct supply of information to the users of the software by the reporters themselves and elements of the use of intermediaries to transmit the information both by requiring (threat of public disclosure) and by enabling (by giving vulnerability information to the vendors) the vendors to disclose the vulnerability information, several actors participate the implementation process. The implementation of vulnerability reporting, the actual process of reporting and its management, is thus only partially at the hands of the discoverers who do the reporting.

The actual process of vulnerability reporting is a combined action of several stakeholders where the procedures, practices and institutions adopted especially by the coordinators between reporters and the vendors (e.g., various CSIRT organisations) play a significant role. The effects of reporting are altered significantly by the different approaches to the reporting process.

---

[650] Economics explain the argument in more general terms. While developing a model where reputational concern of the long-run player to look good in the current period results in the loss of all surplus, Ely and Välimäki, Bad reputation, p. 785-814, show that in models where all parties have long-run objectives, such losses can be avoided.

[651] There already are technological solutions in the software tools that are used to manage vulnerability reporting process that make this information easily available.

The approaches of specific organisations and its stakeholders have been codified and made explicit in the policies and guidelines put forward. In addition to the discoverers/reporters (individuals or organisations that find the vulnerabilities), such explicit policies have been presented also by the vendors (parties that develop or maintain software products that may be vulnerable) and the coordinators of the communication. Without going deeper into the recognised communication problems between the actors in vulnerability reporting[652], the influence of the implementation of vulnerability reporting on the effects it can produce and the possibilities that it has for affecting secure software development can best be analysed by concentrating on the disclosure policies and guidelines that try to formalise the process[653]. From them the different approaches to vulnerability reporting can be seen. In addition to the influence of the differing approaches, the influence of the common and similar points in the implementation of vulnerability reporting can be analysed especially from guidelines and best practices that try to find common ground[654].

---

[652] The communication and its problems have been analysed by Tiina Havana in an empirical study. The results are reported in Havana, *Communication in the Software Vulnerability Reporting Process*, and Havana and Röning, Communication in the Software Vulnerability Process. These communication problems are the actual reasons for the need for the codification of the practices and procedures followed by organisations into the form explicit policies.

[653] Their purpose is to ease the cooperation of and communication between the different stakeholders (discoverers/reporters, vendors, coordinators, and users) in the process of reporting and, in the end, fixing of security vulnerabilities. This is visible especially in the guidelines searching for a common solution, such as the OIS *Guidelines for Security Vulnerability Reporting and Response*, p. 4. Note however, that The OIS guidelines concentrate on the relationship between the discoverer/reporter and the vendor. No claim of universal applicability to all stakeholders is made.

[654] The disclosure policies and especially the guidelines and best-practices that try to find common ground can also be seen as a separate process-based regulatory instrument as such and could be analysed on the basis of their objectives, substance, implementation and reactions similar to any other instrument. However, since their main purpose is to formalise and ease the communication between the actors in the vulnerability reporting process, in

The reason for various approaches adopted in the policies, the rules and procedures set, is that when the policies are drafted from only one perspective (be that the vendors, the discoverers, the coordinators), the interests represented differ[655]. As already noted, even though both vendors and reporters state that security is the most important value in the reporting process, the reasons for the interest in security are still different. Whereas the reporters seem to strive for a level of security that is beneficial for the general public, the vendors are interested in the level of security that fulfils only their customers' needs.

The practices and processes adopted and formalised into policies by the coordinators represent the objectives of the respective organisation and typically serve their internal needs. For example, CSIRT teams in big companies or in academic networks typically coordinate action only inside the member organisations and the public interest is not their main concern. Naturally other teams do not explicitly oppose to this, but they seem to serve more limited purposes and interests.

Of the policies set up by coordinators, the objective of enhancing the public interest in information security and to aid to reach a level of security that is sufficient for the whole society is most clearly protected by the national governmental teams[656]. National security

---

this study they are only considered from the implementation perspective of the reporting as such. A separate regulatory analysis is left for further study.

[655] The policies are typically drafted from the perspective of different organisations as noted by Laakso et al. in Introducing Constructive Vulnerability Disclosures. A list of different policies and guidelines can be found from the web pages of the Oulu University Secure Programming Group (OUSPG) http://www.ee.oulu.fi/research/ouspg/sage/disclosure-tracking/#h-ref10 [updated 10.5.2005, visited 22.2.2006].

[656] For example, the main function of CERT-FI is to promote security of the information society by assisting its clients (every Finnish private individual, company or representative of the government) in preventive actions for computer security incident and the minimisation of the risk of vulnerability. CERT-FI also receives the telecommunications operators' reports on information security incidents and threats demanded by Section 21 titled "information security notifications" of the Finnish Act on the Protection of Privacy in Electronic Communications (516/2004). An unofficial translation of the Act made 1.1.2005 is available via Finlex at http://www.finlex.fi/fi/laki/kaannokset/2004/20040516 [22.2.2006]. See version 1.1 of the charter of

> interests, such as the protection the critical infrastructure, are also high in the list of objectives of national teams[657]. This raises the importance of the disclosure policies advocated by national coordinators (like US-CERT, Aus-Cert, etc.), since they posses the ability to use the policy as a leverage to modify the incentives other stakeholders face for the benefit of the society at large[658].

Conflicting objectives and consequent differing approaches has been proposed to be standardised in order to make the reporting more efficient and effective[659]. One of the first proposals was a draft Internet standard published in the form of a request for comment (RFC) in the auspice of the Internet Engineering Task Force (IETF).

---

the CERT-FI headings 3.1-3.4, published 7.6.2004, available in pdf -format at, http://www.ficora.fi/suomi/tietoturva/certtoiminta.htm (in Finnish) [updated 26.10.2004, visited 22.2.2006]. The charter is based on Internet Engineering Task Force's best practice as presented by Brownlee and Guttman, *Expectations of Computer Security Incident Response*, in RFC 2350.

[657] See, for example, the web pages of the US-CERT, http://www.uscert.gov/aboutus.html [updated 21.2.2006, visited 22.2.2006].

[658] Due to their role in enhancing the public interest, the national coordinating teams are in an especially good situation to develop disclosure policies that minimize total social costs, and not just private costs of a specific stakeholder. Cooperation at an international level is a minimum requirement for this to happen. The priority of governmental funded coordination under welfare economic considerations is shown by Kannan and Telang in An Economic Analysis of Market for Software Vulnerabilities.

[659] The need for codified and explicit practices and processes is clear. The failures in the reporting process, which typically result from failures in the communication between the stakeholders, pose risks for the information security of individuals and organizations. For example, if a publication of a vulnerability is mistimed or information leaks from the parties, the users face increased vulnerability. The same happens if vulnerabilities are not being fixed due to misplaced or –handled reports. The positive effects on secure software development that vulnerability reporting can produce depend heavily on the success of the communication process. This requires clear and explicit rules and procedures in order to be able to balance the differing interests and objectives, and, at the same time, to capture to benefits of reporting and minimizing the risks posed. Even though reporting policies in general try to do this, the consensus seeking guidelines are more effective due to the rules and procedures being standardised and accepted by the majority of stakeholders.

The proposed responsible disclosure draft got rejected due to IETF not considering itself as proper venue for non-technical procedural standards[660]. Most prevalent of current proposals are the OIS (Organization for Internet Safety) *Guidelines for Security Vulnerability Reporting and Response*[661] and the NIAC (National Infrastructure Advisory Council of the U.S.) Guidelines that are presented in the final report of its Vulnerability Disclosure Working Group drafted by Chambers and Thompson, *Vulnerability Disclosure Framework*[662].

The effort is to bring the different actors in vulnerability reporting together to form a common approach that would be applicable to all stakeholders. For example, the goal of the OIS Guidelines and the NIAC Framework is to achieve a common understanding and to develop standard practices for vulnerability reporting[663]. They are designed to eliminate confusion among all stakeholders and the public regarding managing and resolving security vulnerabilities. The coordinating guidelines have the objective of finding solutions that are best possible for all of the stakeholders and that could be accepted by all.

The NIAC guidelines stated objective is to develop a framework for the management of vulnerability reporting in the purpose of advancing national interests. NIAC is a component of the U.S. Department of Homeland Security, and it functions as an advisor of the president of the U.S. in issues relating to the security of information systems for critical infrastructure. However, the guidelines are intended to be non-US-centric with global applicability to all

---

[660] This explanation is provided at the web pages of the Organisation for Internet Safety (OIS), http://www.oisafety.org/about.html#14 [updated 17.9. 2004, visited 22.2.2005].

[661] The OIS guidelines received support from the U.S. National Cyber Security Partnership (NCSP) in 2004, whose task force on the Security Across the Software Development Life Cycle recommended its usage in its report on *Improving Security Across the Software Development LifeCycle*.

[662] See Chambers and Thompson, *Vulnerability Disclosure Framework*.

[663] Note that the acceptance of both the OIS and NIAC guidelines by the security community remains to be seen due to their relatively recent proposition.

stakeholders[664]. In addition, the NIAC guidelines build on existing practices and policies and uses input from industry, government, and academia in the form of literary sources and personal contributions[665]. However, the U.S. concentricity of the guidelines and the emphasis of the national security interests of the U.S. may hinder their worldwide acceptance.

The OIS is explicitly seeking for the public benefit and commit to the public review and commenting process[666]. However, this public interest statement may be diluted by narrow representation only of key commercial software development and security companies[667] and the concentration only on the relationship between the discoverer/reporter and the vendor. No claim of universal applicability to all stakeholders is made. Neither the customers nor the coordinators are represented in OIS[668]. At least there is a threat of other stakeholders trust to the guidelines commitment to solving social problems outside the private interests of the companies being lower due to the representative structure and the concentration being purely

---

[664] Voluntary implementation of the guidelines worldwide is deemed to help in minimising also risks the vulnerabilities pose to the US. Chambers and Thompson, *Vulnerability Disclosure Framework*, p. 12-13.

[665] As visible in the report drafted by Chambers and Thompson, *Vulnerability Disclosure Framework*, p. 12, the sources (literary and personal combined) represent all of the stakeholders in vulnerability reporting; the discovers of the vulnerabilities, the vendors of the vulnerable software, the users of the software, and the coordinators of the vulnerability reports.

[666] See the web pages of the OIS, http://www.oisafety.org/about.html#4 [updated 17.9.2004, visited 22.2.2005]. The wide adoption and effectiveness of the guidelines is considered to be dependent on their alignment with the consensus of the security community.

[667] Only big software development and security companies are represented in the OIS at the time of writing. See http://www.oisafety.org/about.html [updated 17.9.2004, visited 22.2.2005] and the list of adopters at http://www.oisafety.org/adopters.html [updated 17.9.2004, visited 22.2.2005].

[668] The OIS is forming an advisory board that would represent the interest of computer users. Recognized security experts have been asked to join. See the web pages of the OIS, http://www.oisafety.org/about.html#15 [updated 17.9.2004, visited 22.2.2005].

on vendor-discoverer relationship[669]. However, the OIS represents such an influential group that can act as gatekeepers for the vulnerability reporting process. The processes that the big software development and security companies adopt are likely to affect the opinions and possibilities of others also.

It is not possible to go into the practical implementation of all the rules and procedures that these policies and guidelines contain and all their possible influences to the reporting process. Only certain central questions and solutions to the main communication problems can be raised.

One of the crucial issues for the effectiveness of vulnerability reporting and a central demand in the policies and guidelines is the establishment of a clear point of contact for vulnerability reports (e.g., a single dedicated email address) by all the stakeholders. This is especially important for the vendors since the reporters seem to have the biggest problems in finding the right persons to contact[670]. Another issue that centrally increases the effectiveness and speed of reporting is that stakeholders provide and clearly indicate (e.g., in web pages) consistent formats for reporting and displaying security vulnerability information, make explicit all the procedures for handling security vulnerabilities including the policy, and that the vendors clearly indicate a list of supported products and their versions.

Proper measure to protect the sensitive vulnerability data (to keep its integrity and confidentiality) and to verify the authenticity and non-repudiation of participants of the communication is also an essential part of implementation. Not only for the purpose of encouraging finders of vulnerabilities to communicate vulnerabilities to software

---

[669] Suspicion of the bias towards vendor preferences is clearly visible in the public comments of the first version of the OIS guidelines. See comments from public review http://www.oisafety.org/reference/comments.pdf [22.2. 2006].

[670] Finding the right contact has been reported to be difficult especially for the reporters by Havana and Röning in Communication in the Software Vulnerability Process, p. 7.

vendors[671] but also in order to gain legitimacy and trust in the eyes of the stakeholders in the reporting process. Due to premature leaks of vulnerability information imposing increasing risks to users, the protection of the information, by all the relevant parties to the vulnerability communication process, is deemed necessary[672]. Clear indication of how the sensitive data is protected is also needed in order to increase the trust of other parties and the public in general. This is especially important for the national coordinators since they operate on the basis of trust from the stakeholders and the public in general[673].

However advisable the use of encryption, for example, PGP and equivalent open source version like Open PGP, and other protection measure is, it can also have a negative effect on the possibility for vulnerability reporting to influence the behaviour of the vendor. Many encryption and digital signature products still do not interoperate well and the costs of maintaining several systems with the same purpose are high. This may lead major actors (mainly vendors or coordinators that handle vast amounts of vulnerability data) attempting to sign on only to a single product. However, this would limit the pool of participants, force them to use other, more inconvenient and costly

---

[671] As reported by Tiina Havana in *Communication in the Software Vulnerability Reporting Process*, p. 41-42, the communication channel most commonly used at both ends is email. Due to the sensitivity of the information reported, encryption is increasingly used especially by the reporters. The use of both encryption to preserve the integrity and confidentiality of message contents, and digital signatures for authentication of the reporters and the non-repudiation of the reports helps the reporter to increase the credibility of the claim. The vendors are more likely to be responsive since they can at least know the identity of the reporter and the integrity of the report.

[672] This is emphasised especially in the wide consensus based NIAC guidelines, where the protection of information is an essential part of the guidelines for all the stakeholders.

[673] As pointed out by Chambers and Thompson in *Vulnerability Disclosure Framework*, p. 34, for vulnerability coordination to be effective, vendors, reporters and users must be aware of the existence and legitimacy of coordinators.

(like traditional mail), secure communications methods, or lead them to stop using secure communications methods altogether[674].

Other rules and procedures that the reporting policies and guidelines typically include concern, for example, what the finder should do before reporting (e.g., check whether the issue has previously been publicly identified and remedied), rules on the vendor notification confirming it has received the report (timeline, usually 7 business days, and what the reporter should do if no confirmation is sent; e.g., send a request for confirmation and turn to conflict resolution), obligations to give status updates to the reporter. Also further rules on giving security advisories are typically involved[675] together with standard forms for the different stages of reporting.

Due to the codification, the information regarding the procedures and practices of vulnerability reporting can be more easily taught and transferred inside organisations. There is no need for everyone to learn the practices and procedures by doing. This alone positively influences the efficacy of vulnerability reporting. When the policies and guidelines are sufficiently clearly made available to the stakeholders, they are made more aware of what others expect from them and understand the role and importance of reporting in the organisation of other stakeholders. This increased understanding of the behaviour of others can further decrease the possibilities for conflicts between the stakeholders.

Note that the implementers of vulnerability reporting, those adapting practices and procedures necessary for the effective functioning of the reporting, have a considerable discretionary power. Not even the guidelines and the policies are coercive. In addition there is no other monitoring than the negative or positive publicity of those

---

[674] This is noted Chambers and Thompson in *Vulnerability Disclosure Framework*, p. 22, which codifies the standpoints of the vulnerability disclosure community at large.

[675] A security advisory provides information to the general public about the security vulnerability, the products and versions it affects, and the steps that can be taken to defend affected systems and networks against it. However, there is no clear consensus on the amount of information that should be provided and when to do so.

either complying with them or not. There neither is external enforcement of the rules outside the ability to turn to third parties like the coordinators or to exit the process and to publish the vulnerability or shut down communication[676]. Actually, the desired commitment to rules and procedures is based on voluntary acceptance due to them providing effective ways to manage and response to vulnerabilities in software. Coercing stakeholders to obey the rules does not belong to this scheme and the non-coercive approach seems effective at least as long as the stakeholders remain motivated to comply with them.

Even if the implementation of vulnerability reporting is based on the consensus seeking guidelines, it does not necessarily lead to similar practical implementations in different organisations[677]. This might be the case despite the separate implementation assistance provided[678]. Implementation may vary on the basis of the different needs and objectives of the organisation. This may partially dilute the objectives of enhanced communication between stakeholders and better vulnerability response sought by the guidelines and necessary for vulnerability reporting as an effective regulatory instrument, especially

---

[676] Rules on conflict resolution between the vendor and the finder are typically also included in the policies. These rules differ, but typically stipulate to possibility to exit the process of following the rules and practices of the policies or the guidelines. For example, the OIS guidelines include rules on exiting the process in the situation of irreconcilable disagreements. However requirements of reasonable efforts to solve the problem and noticing other parties before exiting are provided together with the possible use a third party to solve the problems before exiting.

[677] The guidelines do not even require this. The OIS guidelines require that when adopting the guidelines as the basis of the organisational vulnerability reporting policy, the policy should note that it is based on the guidelines provided by OIS and clearly note every deviation from them (OIS, *Guidelines for Security Vulnerability Reporting and Response*, p. 2). The NIAC guidelines state that stakeholders need to choose whatever actions are appropriate for their circumstances and environments (Chambers and Thompson, *Vulnerability Disclosure Framework*, p. 27).

[678] OIS publishes a document that helps to implement the guidelines for security vulnerability reporting and response. Current version 2.0 is available at http://www.oisafety.org/reference/implement.pdf [22.2.2006].

if the policy maker considers only its own objectives and interests while deviating from the guidelines.

The formalisation of the procedures for vulnerability reporting can also have negative effects. The complexity, length and excessive formality of the guidelines and the consequent policies might expel those reporters that do not have large enough resources to vulnerability research. Persons that are not deeply involved with the reporting are not likely to study and comply with the complicated and lengthy policies, especially if they are not likely to find many vulnerabilities in their life-time. This means that these vulnerabilities are not reported at all, or reported by using other, more endangering or less effective channels. The length and complexity of the policies might diminish their effectiveness.[679]

A similar negative effect can be caused by the policies and guidelines bias in favour of the vendors. If the reporters do not consider their interests being represented by the policies, they are less likely to comply with them and report accordingly. There still are significant areas of dispute in the guidelines.

It is not clear whether the reporters are going to be less active or even more active than currently with the wider usage of policies, especially those basing on common guidelines. The guidelines do not, similar to the existing policies, provide other than reputational rewards for the reporters[680]. If media attention fades due to the reporting becoming part of normal development work and seeming a less radical action, at least part of the attention and career enhancement possibilities seeking reporters are likely to report less[681].

---

[679] This has been pointed out in the comments for the first version of the OIS guidelines. See OIS web pages; section Summary of Changes from Public Review, http://www.oisafety.org/changes.html [updated 17.9.2004, visited 22.2.2006].

[680] Typically the policies and the guidelines expect the name of the reporter to be published, if not denied by the reporter, in the advisories provided to the public.

[681] However, this is not likely to be a large population since most of the reputational rewards are important only inside the professional group. Wide media attention in public press is not needed for reputation to increase.

If the consensus based guidelines are taken into wide use, they can provide a standard of care in some jurisdictions. Enforcing social customs or best practices in an industry is one of the ways legal standards can be set. It can develop into a metrics towards which the actions of the stakeholders are compared against. It can become the basis for the analysis of liability under tort law – no further evidence of the basis for action is needed; either the vendor or the reporter could be held negligent due to the failure to adhere to the consensus showing practices and procedures in a specific case (assuming occurred damage and causation are proved). Doing this in an early stage of best practice development without the court actually balancing marginal costs and benefits can lead to inefficient norms being upheld and the vulnerability management being hampered[682]. However, this has not been done and it is not clear, and by no means obvious, that courts will do so. If it is done, it would raise the implementation of vulnerability reporting a level higher and set a legal liability for proper management of vulnerability reports.

A severe hindrance for cooperation in the vulnerability reporting process and especially for efficient communication between the different stakeholders is posed by differing understandings of the severity of the threat posed by a vulnerability. When a vulnerability is reported it typically is scored, i.e., assigned a metric that communicates a sense of the severity of the vulnerability. This score is used as a basis for assigning importance to the remediation efforts and it influences also research and publishing efforts of stakeholders. For example, if the reporter sees the threat posed by a vulnerability as more severe than the vendor does, this disagreement might lead to vendor being reluctant to issue patches or to provide information of the vulnerability to the users and, at least, problems in establishing a common timeframe for patch and information release between the

---

[682] Since the lawmaker or the court in this type of standard setting relies on the community of people who created the norm or the industry that engages in the practice to balance costs and benefits, it ought to be made ascertain whether the balancing is actually done. This is discussed by Cooter and Ulen in *Law and Economics*, p. 315-316. Enforcing community standards where all the stakeholders are not even represented could turn out to be counter-productive.

vendor and the reporter are created. Then it might be necessary to bring in a coordinator to assess the risk posed and the communication might even collapse altogether.

Currently there are several ways of scoring the severity of a threat that a vulnerability poses in use and disparate methods seem to yield different results[683]. This negatively influences the resolving of a vulnerability because, to the extent different scoring methods are adopted by the stakeholders, a basis for disagreement is created. As noted in the Chambers and Thompson "[t]he weaknesses or dangers associated with a vulnerability may be exacerbated by those disagreements, and provide malicious actors increased time to exploit the vulnerability or increase the damages resulting from existing exploitative situations"[684].

A central question in the implementation of vulnerability reporting is the establishment of specific organisations for vulnerability management and coordination of reporting[685]. The coordination

---

[683] This is a result of the evaluation of the Vulnerability Disclosure Working Group of several alternative scoring methods actively employed by stakeholders to categorize reported vulnerabilities. They claimed to have run several past worms, viruses, and software vulnerabilities through existing scoring methods and the results are stated to differ wildly. See the Meeting Minutes of the National Infrastructure Advisory Council (NIAC) meeting for 14 October 2003, p. 18, at http://www.dhs.gov/interweb/assetlibrary/NIAC_Final_Minutes_101403.pdf [22.2.2006].

[684] Chambers and Thompson, *Vulnerability Disclosure Framework*, p. 21-22. The need for a common vulnerability scoring method has been raised and the NIAC is addressing this problem in a specific scoring subgroup of the Vulnerability Disclosure Working Group that is supposed to publish a separate report. See NIAC Vulnerability Disclosure Framework Report, p. 8.

[685] The CERT Coordination Center (CERT/CC), the first organisation to coordinate communication among experts during information security emergencies, grew from a small computer security incident response team formed at the Software Engineering Institute at Carnegie Mellon University (Pittsburgh, Pennsylvania, USA) established by DARPA (Defence Advanced Research Projects Agency) in 1988. Currently CERT/CC http://www.cert.org) is a non-academic and non-governmental unit funded mainly by the U.S. Department of Defence and the Department of Homeland Security along

centres, also the national CERT/CSIRT teams, operate differently in each organisation and member state. There is a variety of institutional forms and backgrounds even though the procedures and practices are based on similar principles and guidelines. This makes the cooperation complex and inevitable decreases the effectiveness of the coordination centres against international problems. A special fear in Europe has been that worldwide coordination is done through partly U.S. government funded CERT/CC and that CERTs in Europe are dependent on its information release policy[686]. As a response, EU is studying ways to strengthen cooperation between EU member state CERTs and CSIRT teams and other similar organisation[687].

A special problem in the implementation of coordination centres, especially at the national level, has been the assignment of sufficient resources. Especially the national teams play such an important function in society (computer emergency response and coordination

---

with number of other federal civil agencies and private sector. It was followed by several other teams that serve as coordinators for various regimes; in addition to governmental teams such as the US-CERT (U.S. governmental CERT, http://www.us-cert.gov), DoD-CERT (U.S. Department of Defence CERT, http://www.cert.mil), Aus-CERT (The Australian CERT, http://www.auscert.org.au), CERT-FI (The Finnish CERT, http://www.cert.fi) set up by government representatives, there are also dedicated teams in commercial and academic organisations like the FUNET-CERT (Finnish University and Research Network – Computer Emergency Response Team, http://www.cert.funet.fi)) and many internal teams in large enterprises, set up by the organisations themselves. CERT/CC has for its part helped to establish these other teams and coordinates with these other teams for incident response, but it remains independent of them in terms of organisation. Many of these teams are members of the Forum of Incident Response and Security Teams (FIRST, http://www.first.org), that has operated as a coalition to exchange information and coordinate response activities among the large variety of specialised groups since it was established 1990. CERT/CC is even a founding member.

[686] This fear has been raised by the European Commission in its communication on network and information security, COM(2001)298, p. 21.

[687] The establishment of the European Network and Information Security Agency (ENISA, http://www.enisa.eu.int/) is part of this. See Regulation (EC) No 460/2004 of the European Parliament and of the Council of 10 March 2004 establishing the European Network and Information Security Agency, Official Journal L 077, 13/03/2004, p. 1–11.

in addition to vulnerability coordination) that they should be adequately equipped. For example, in Finland the tasks and responsibilities of the CERT-FI group operating in the Finnish Communications Regulatory Authority (FICORA) have been widened several times, but the resources directed have not increased with similar speed. Perhaps this speaks about the importance given to the problem of vulnerability reporting and incident response in the government. At least it means that the possibilities of CERT-FI group are still not fully taken into action and it has for long largely remained as a filter of vulnerability information coming from other coordination organisations.

Note that when establishing a computer emergency response team (CERT) or a computer security incident response team (CSIRT), the objective typically is not solely on vulnerability reporting. The main purpose served by these teams, no matter what organisation establishes them, is to help in responding in situations of attacks and other incidents on computer systems. Vulnerability management and coordination of reporting is typically only a secondary objective, even though an important one[688]. According to the common approach of the stakeholders in the vulnerability reporting process, even when established for vulnerability management and coordination purposes, the coordinating organisations serve several objectives depending on the organisation and its needs[689].

---

[688] This is visible in the widely used basis for the ground rules both for CERT and CSIRT teams IETF RFC 2350, Brownlee and Guttman, Expectations for Computer Security Incident Response, p. 11.

[689] Chambers and Thompson, *Vulnerability Disclosure Framework*, Appendix B, p. 44-45. *Research institutions* may establish vulnerability coordination activities with the objective of supporting or promoting their research agenda and findings. A common subcategory of this type is that a research group may engage in vulnerability coordination for issues it discovers. *Government organisations* objective in establishing vulnerability coordination centres is to serve the public interest and to protect the critical infrastructure and national security. This objective is typically visible also in the policies of the national coordinators. The objective of *commercial organisations* in establishing vulnerability coordination activities is to help in corporate governance and information security. The objective is to serve the internal needs of the company and to internally

Despite the secondary role of the coordination of vulnerability reporting in the activities of these organisations, they centrally influence the reporting process. Their role as a coordinator is to supervise the work related to vulnerabilities by assisting other participators and gathering knowledge in one place. They assist in the disclosure and response to new vulnerabilities. Basically the coordinator forms a communication link between the reporter and the vendor. They provide contacts; function as a wide known contact point, and as an additional medium through which to communicate with the public or with multiple vendors[690]. Coordinators also provide major source of vulnerability information for the vendors. Also a dispute resolution in the form of independent evaluations of both vendor and reporter claims can be part of their repertoire.

Coordinators main function in vulnerability reporting is to help in the communication between the reporters and the vendors. In addition, some of CERT like organisations conduct vulnerability analysis themselves and verify of the accuracy vulnerability reports. However, this is not common. As visible in the guidelines for vulnerability reporting, mediators are typically called into action only when the reporter and the vendor cannot reach a consensus.

Coordinators are often also vested with functions that benefit pre-release secure software development more directly. They can study relationships between vulnerabilities and their reasons, and recognize

---

manage the communication on vulnerabilities. These organisations typically do not provide coordination services between and among vendors and are usually not involved in predisclosure activity. However, they may serve as the vehicle by which specific corporate security concerns are addressed with software vendors. In many cases, these organizations have an assessment and compliance role as well. *Security solution providers* objective in establishing vulnerability coordination activities often is to create a way to ensure their products are up to date.

[690] According to Chambers and Thompson, *Vulnerability Disclosure Framework*, Appendix B, p. 44, coordinator acting for a small constituency may be in the best position to evaluate threats to that constituency and provide installation-specific advice and guidance. Coordinator acting on behalf of a larger constituency often has the ability to draw a great deal of attention to problems affecting a large number of users.

larger trends in vulnerabilities. The research into the nature and causes of vulnerabilities that may be shared with leading research groups can help in the education of secure development methods and practices outside the affected parties.[691]

Coordinators generally do not act with direct authority over a product, not even teams that serve purely internal organisational purposes, and in most cases do not act with legal authority to compel a vendor or discoverer toward any particular behaviour. Not even the national coordination teams typically posses such powers. However, this lack of coercive power does not limit their effectiveness, quite to the contrary. Their effectiveness is based on the trust placed on them by the different stakeholders. Most groups acting as coordinators today have achieved their legitimacy through demonstrated effectiveness, having established a broad audience of system administrators and end users, and by having established productive working relationships with software vendors. In many cases, they have demonstrated their effectiveness for their audience and gained legitimization.[692]

However, the different institutional backgrounds and purposes served by coordinators, even of national teams, highlight the need to implement the correct organisational structures, checks and balances, and decision-making rules in order to establish a sufficient amount of independence from the organisation in question or the government and the customers. The legitimacy and trust needed for the coordinators to work effectively require that. The problem of capture, a situation in which the coordinator is embedded in such close relationships with a specific stakeholder that it sees its functions as that of safeguarding the interest of that stakeholder above all might exists in teams that serve only the interest of a specific stakeholder (e.g., a research group or a vendor).

This typically is not a widespread problem. It could be if a strong enough stakeholder could compel others to follow its policies and

---

[691] This has been pointed out by Chambers and Thompson in *Vulnerability Disclosure Framework*, Appendix B, p. 44.

[692] Chambers and Thompson, *Vulnerability Disclosure Framework*, Appendix B, p. 45.

practices. In the case of national teams, serving the public interest, the threat of capture is relieved by them not only regulating one industry or interest group (e.g., the vendors) but, instead, acting as a mediator between different stakeholders. They need to gain the legitimacy of all stakeholders in order to effective.

National governmental teams acting as coordinators and being funded by the government are in a unique position to compel others to follow their policies that typically are more biased on the public interest than other policies. However, this is not typically done and the policies for vulnerability reporting concern only those who utilise the national coordinating teams in their reporting. National CERTs and alike do not use their regulatory capacity in the public interest at the level of policies[693]. Usually there is no obligation to invoke a coordinator in a reporting process; direct contact between the vendor, the discoverer, and the users of the product is also widely used.

*Lower level of package.* The way the vendors notify their customers about vulnerabilities change the final effects of vulnerability reporting can have on the behaviour of the users and the general public. The first issue in the implementation by the vendors is their expertise and interest in handling vulnerability reports. Previously, as already discussed, the vendors as receivers of the vulnerability reports have had a negative attitude and responded accordingly. They have

---

[693] The public interest is sought by means of informing and assistance, not by developing socially optimal reporting policies. Even though the national teams are in the best position to enforce socially optimal reporting policies due to their role in advancing the public interest, it is not clear that they could or even should do so. Since they act as mediators in the process, their policies would only be complied widely (assuming that the socially optimal reporting policy is not the one that gains wide acceptance by all stakeholders) if there would be an obligation to invoke a coordinator in the reporting process. Since the role of national CSIRT teams is just to assist and advance the communication and cooperation between the main actors (the vendors and the reporters), this might be desirable only if the vendors and the reporters could not cooperate in a socially optimal way.

understated the problems or denied them altogether, tried to keep things secret, tried to deny liability or to shift blame[694].

However, categorical denials of the existence of a problem or the threats of legal attacks against reporters have become damaging to the vendors reputation. This is due to the increased likelihood of the problems being confirmed later on and threats of legal attacks reaching high media attention and public outcries[695]. Research, awareness and interest in security vulnerability issues simply has raised and the media is especially attentive to these issues.

The manner in which commercial software product development companies respond to security issues and especially to vulnerability reports has evolved due to the threat of public disclosure accompanied with the increased awareness and a greater concern for vulnerabilities among customers and other stakeholders[696]. Formal processes to discover, evaluate and fix vulnerabilities when they arise are put in place. The policies are useful in this by codifying information and as a way to improve procedural knowledge in an organisation[697].

The speed at which patches are issued or vulnerabilities otherwise resolved (e.g., by issuing workarounds, recognising that no vulnerability exists or has already been fixed) after vulnerabilities are reported also alters the effects of disclosure. The quicker a patch is issued, the sooner the users are enabled to patch their systems and

---

[694] Lack of expertise and interest especially on the side of the vendors was one of the reasons why the vulnerability disclosure became such a heated debate in the information security discussion at the turn of the century.

[695] The possibility of negative effects on vendor behaviour has also been raised by Chambers and Thompson in *Vulnerability Disclosure Framework*, p. 24.

[696] Vast majority of the receivers of the reports (about 70%) stated in 2002, when asked about their general opinion on the importance of bug reports in a quantitative survey conducted by Tiina Havana, *Communication in the Software Vulnerability Reporting Process*, p. 45-46, that there probably are security bugs in their products, and they are important to be repaired for which reason bug reports are of great importance. Almost 90% see that the security bugs existing in their software are important to be identified and that bug reports are important in this (not marginal).

[697] See Havana and Röning, Communication in the Software Vulnerability Process, p. 8.

to minimise the window of vulnerability. This has improved dramatically. At least those vendors who are aware of the vulnerability discussion have become very responsive[698]. Preliminary empirical evidence shows, however, that the size and market share of the vendors influence their responsiveness; large vendors patch faster then smaller ones[699].

The medium by which patches are released also has a potential to alter the effects of vulnerability disclosure. Disclosure by the vendor, typically to the customers in the form of putting the patches available to the web pages or mailing them to the registered users (who have consented to such mailings), might not reach all users. Making information available on the Internet and mailing patches to the users who have consented to such postings may be an easy way to reach certain segments of the users, but not at all promising for reaching people with less familiarity with the vulnerability issues or even with the use of computers like in the case of many home users.

Also the information accompanied with the release of a patch or a workaround can alter the effects. Mere patch release neither might sufficiently raise the interest of all users as such nor encourage users

---

[698] According to the first empirical analysis of the communication in the vulnerability reporting process conducted by Tiina Havana, *Communication in the Software Vulnerability Reporting Process*, p. 74, majority of vendors seem to put priority into vulnerability correction. They at least state, the possibility of reputation management by the vendors is raised, to interrupt other work immediately when they receive a report and concentrate on the repairing process, or have formed a specific schedule within which the reports are supposed to be handled. Note that those answering questionnaires on their vulnerability reporting processes are likely to be more knowledgeable of the issue and likely to be more responsive.

[699] This result has been reported by Arora et al. in Impact of Vulnerability Disclosure and Patch Availability, p. 18. This can be due both to the higher awareness among larger vendors and higher customer demand accompanied with possibilities to direct resources to handle reports. Negative effects on reputation, as proposed in the empirical analysis (idem p. 18), can be higher to the larger vendors but they are also in a better position to cope from it and are less likely to lose the whole business due to the larger and varying customer base. Negative publicity can be devastating for a small start-up company trying to establish its products into the market.

to install patches to their own systems. Installing a patch into a users system typically involves costs and requires time, for example, in terms of testing and re-configuration. The users need to know the seriousness of the vulnerability, its possible consequences, whether is affects their systems, and the possibilities to fix them, in order to determine whether and which patches to install and when. In short, they need to know why they should install the patches. Provision of security advisories that provide information to the general public about the security vulnerability, the products and versions it affects, and the steps that can be taken to defend affected systems and networks against it play an important role in this[700].

Also the manner in which the information is provided may shape the effects. When vulnerability information is not disclosed in clear and usable form, it may actually make people less knowledgeable than they were before. This may produce over-reactions or under-reactions, based on an inability to understand what the information actually means. This emphasises the importance of enhancing the disclosure process.

Same problem with the publication exist as with the direct reporting from the discoverers to the users; in addition to the positive effects of increased awareness of security issues and the possibility to decrease risks by installing the patches, the publication can also increase the risk of attack. Preliminary empirical evidence speaks in favour of increased attack frequency due to the disclosure of a vulnerability, even when a patch is available[701]. The release of a patch for a known vulnerability decreases the number of attacks but the release of a patch for a hitherto unknown vulnerability increases the number of attacks. A suggested explanation to this is that attackers believe that not all

---

[700] Note that security advisories can be published also by other actors in the vulnerability reporting process than vendors. Especially if the vendor does not do this for what ever purposes, the reporter's possibility to publish their own advisories has been recognised. Also the coordinators can publish their own advisories, even though they typically just adjust the already published advisories to their own contexts and republish them. This has been acknowledged by Chambers and Thompson in *Vulnerability Disclosure Framework*, p. 30-31 and 36, and in the OIS *Guidelines for Security Vulnerability Reporting and Response*, p. 20-23.

[701] See Arora et al., Impact of Vulnerability Disclosure and Patch Availability, p. 19.

users will patch in time. The installation of patches into the user systems is one of the pressing issues since majority of attacks take place after the patch has been issued by the vendor but has not yet been widely installed by the users[702].

A crucial issue for the effects on secure software development outside post-release vulnerability correction is whether the information gathered about discovered bugs is used to prevent similar bugs in the future. Most direct influence vulnerability reporting can have on initial security development depends on vendors internalising the knowledge of vulnerabilities, their causes and fixes, gathered in the reporting process and the fixes. Knowledge of a vulnerability, not just of its existence but also of its details (like the specific coding or implementation error that causes it) can be used to train current and educate future software developers more widely to avoid the causes of the vulnerabilities.

The problem is that secure software development practices still largely are tacit know-how that is difficult to transfer[703]. It has to be learnt in contact with the possessor of the knowledge. Only to the degree the knowledge is codified can it be taught. Codification of vulnerability information could serve as an important educational resource and ease the current lack of teaching material. For example, recent textbooks on secure software development, such as Viega and McGraw, *Building Secure Software* or Graff and van Wyk, *Secure Coding*, use known vulnerabilities especially as examples to clarify and make things more concrete

However, the gathered vulnerability information is rarely used in practice for the benefit of building secure software from the beginning (i.e., prior to release). Only about a half of receivers of the reports (vendors) stated in 2002 that they distribute the information about discovered defects to their software developers and make that

---

[702] This will be discussed below in heading 4.3.4 concerning the reactions of objects.

[703] This has been pointed out, e.g., by Lundvall in Understanding the Role of Education in the Learning Economy, p. 18-21.

information an essential part in the software development process[704]. The gathered knowledge is not widely used in the purpose of correcting the pressing problems of information security; the avoidance of similar, perhaps even widely known vulnerabilities appearing over and over again.

Neither do the policies and guidelines typically address this issue. The policies and the guidelines for vulnerability reporting do not provide any rules, practices, procedures, or demands for the information communicated to be used for the benefit of initial secure software development. For example, there typically are no demands for the vendors to pass the information about discovered vulnerabilities to their software developers in order to prevent similar vulnerabilities in the future. Actually the guidelines do not even emphasise the introduction of a vulnerability into a product and the life-cycle of a vulnerability depicted in them begins with discovery of the existence of the it in a specific software product[705].

---

[704] Havana, *Communication in the Software Vulnerability Reporting Process*, p. 69-70. Whereas little over half of the receivers of the reports answered that they pass the information further to their software developers with the intention of preventing similar vulnerabilities occurring again and only 15 per cent of all the respondents to the questionnaire do this, and when they do it, the information does not have an essential role in the software development process. Havana and Röning, Communication in the Software Vulnerability Process, p. 8.

[705] The formal life-cycle model of vulnerabilities presented both in the OIS *Guidelines for Security Vulnerability Reporting and Response*, p. 2-3, and in the recommendation for NIAC guidelines presented by Chambers and Thompson in *Vulnerability Disclosure Framework*, p. 13-16, begins with discovery. However, the NIAC guideline at least acknowledges the introduction of the vulnerability into the product and then starts with the resolution of the vulnerability. Even though this could be explained by the shorter writing style of the OIS guidelines, it still represents the differences in attitudes and objectives. The major software vendors and security companies behind the OIS guidelines may not want to admit the existence of their liability for the vulnerabilities found from their software. This could be interpreted as a desire of the vendors to use vulnerability handling guidelines to express in a subtle manner not only the inevitability of vulnerabilities but also the limited responsibility of the vendor to develop initially secure software products. At least, this is the image that the guidelines easily gives to outsiders.

Basically the reporting of vulnerability information could also contribute to improving the engineering quality of software products, by supporting the academic and research communities' ongoing efforts to identify common security vulnerabilities, the conditions under which they occur, and methods to avoid them[706]. However, the policies and the guidelines provide no procedural support for this and the concentration is purely on fixing the problems that occur after release.

This further diminishes the possibilities of vulnerability reporting to affect initial secure software development and shows an issue that ought to be altered in order to reach more effective influence. If vulnerability reporting does not shift from customer management to an essential source of information for software development in the developing organisations, the effects of vulnerability reporting on initial secure software development is minimal.

### 4.3.4 Reactions of objects

As discussed above, vulnerability disclosure as a regulatory instrument may influence behaviour in two ways; through decisions by users to alter their behaviour with respect to the risks involved (install patches or take other protective measures) or through decisions by vendors to alter their products in ways that make them less vulnerable. Knowledge of the vulnerability of a software product not only enables the protection measures to be taken in order to shield oneself from the harms or to make more rational purchasing decisions, but also incites software vendors to increase the level of security in their products. Basically both of these influence mechanisms are affected by the reactions of the objects of regulation, i.e., the (potential) users of software.

Basically, an assumption of a self-interested and fully-informed rational actor that can react to the information provided to him underlays the influence mechanism. In reality, however, the users

---

[706] This is made explicit in the OIS *Guidelines for Security Vulnerability Reporting and Response*, p. 1.

especially have limited time, energy, and attention, and therefore, may have limited ability or willingness to properly process and act upon information, even assuming perfect accuracy and effective dissemination. Moreover, the inherent complexity and significant uncertainty that often characterise vulnerability information substantially impedes the ability to communicate information accurately, effectively, and in a manner that does not lend itself to distortion. Even if such communication problems can be overcome, there is little empirical evidence available to allow more than theoretical speculation on how individuals will act upon information conveyed and how such action specifically influences market and social behaviour. Due to the lack of evidence, which calls for further study, only few speculations can be made.

The effects depend to a degree on the expertise and level of knowledge of the objects of regulation. This can vary significantly between home users and organisational users that can be end-users (i.e., individuals using software in an organisation) or those purchasing and implementing the software product into the organisations information system. In cases where the users are traditional consumers that are not knowledgeable of the vulnerability issues and do not see the importance of vulnerabilities outside their own system (e.g., when a consumer's system is involuntarily and even unknowingly used to launch a DDoS attacks or to send spam) they are likely to dismiss the information provided by the reporters, vendors, or coordinators, as irrelevant. Due to their relative lack of interest in security issues, the likelihood of patches being installed in a timely manner is low.

Knowledge of the vulnerabilities, even when combined with the above mentioned information of their severity and of the affected systems in the form of security advisories as such does not necessarily lead to patches being installed or other protective measures being taken even by the organisational actors. For example, due to the current high amount of fixes being issued to single widely used software product, which there can be several inside an organisation's information system, in an increasingly rapid pace means that keeping updated is a highly cumbersome and resource consuming task. If patch

installation and management is not sufficiently addressed to and made easier, the users are likely to be overwhelmed by the mere number of patches to install into various products[707].

In addition, as anecdotal evidence from practitioners explicate, most people upgrade for newer functionality, the hope of more robust software, or better performance, and not just because they know of a real vulnerability[708]. This is why it takes a long time before most people upgrade to patched versions[709]. Users are subject to prolonged risk due to this, since the window of vulnerability stays open[710].

The influence of the reliability of the information source for the actions taken by the vendors was already pointed out above. Similar effects can be seen also in the reactions of the final objects of

---

[707] Patch automation at the implementation level, i.e., automation of the installation of the issued patches by the vendor, is one of the methods to make their management easier for the users. However, there is some reluctance to its use due to the consequent problems after patch installation. The least of these problems are not the side-effects produced by patch installation to the operation of the whole information system. The testing needed to see the compatibility problems to the specific circumstances of large information systems often prevent automated patch installation. However, this is something that can be enhanced at the implementation level and will probably diminish when the automation matures and the operators of large systems continue to test for themselves.

[708] Viega and McGraw, *Building Secure Software*, p. 16.

[709] According to Arbaugh et al., Windows of Vulnerability, p. 52-59, there are at least three reasons why not all users patch the minute the patch is released. First, it takes time to disseminate the patching information to all users. Second, some customers lack the requisite computer skills. Third, some users are aware of the patch, but would wait to be sure that the patch is more likely to prevent damage than it may cause.

[710] In Windows of Vulnerability Arbaugh et al., p. 58, in discussing a life-cycle model for system vulnerabilities, made the conclusion that almost all reported intrusions could have been prevented had the systems been actively managed, with all security-relevant corrections installed. However, because the installation takes time, exploits continue to occur even after a patch is issued (sometimes years after). In their model, intrusions increase once a vulnerability is discovered, the rate of intrusions continues to increase until the vendor releases a patch, but the exploits continue to occur in slowly diminishing rate even after the patch is issued.

regulation, the users. With more reliable sources of vulnerability information, the users are more likely to react accordingly and to install the patches provided or to take other protective measures. Vulnerability information is distributed by several sources, the original source being the reporter. However, in the eyes of the users they are likely to be less reliable sources than the vendors that have verified the existence of a vulnerability and provide an accompanying patch. Vendors are more likely to be seen as reliable source of information in the eyes of the customers because the information source is known and additional measures like signatures can be used to authenticate the disclosure. Even more reliable sources of vulnerability information, at least in the eyes of individual users, are the national CSIRT teams or other official public coordinators like the CERT/CC.

The level of details of the vulnerability published has effects on the objects of regulation. Making detailed information explicitly available to the wider public has conflicting consequences because the information is useful both in correcting and exploiting vulnerabilities. The details give precise information for the vendor to verify the claim and to create a patch, and for the user of the software to verify that the vulnerability exists in her systems and that a patch is efficacious. However, if the details become more widely known, e.g., in the case of disclosure via full disclosure list like the Bugtrack, also the attacker is helped at the same time in the creation of exploits[711], in the determination of vulnerable systems, and in the launching of attacks[712].

---

[711] According to the participants in the software vulnerability reporting process, with full details it is relatively easy for fairly unskilled malicious individuals to develop exploits thus increasing the immediate risk to end-users. See, Chambers and Thompson, *Vulnerability Disclosure Framework*, p. 30.

[712] Script Kiddies are also using the information and this greatly increases the number of potential attackers. Note that more advanced crackers can misuse an unknown defect longer when others do not know to protect from that vulnerability. When a good guy finds the defect and publishes it she, at the same time, increases the number of potential attackers with less skills (and less remarkable effects; is it then just an annoyance?) but also makes it possible for the system operators to fix their machines.

Without going deeper in to the effects of explicit details (e.g., proof of concept code at source level) it can be stated that automation of a vulnerability (creation of an exploit), not just its disclosure, serves as the catalyst for widespread intrusions[713]. While the explicit details serve both the correction vulnerabilities, their exploitation and the defence of information systems, the automation of vulnerability exploits seems to serve only the exploitation of vulnerabilities.

Even though the inherent limitation of the substance of vulnerability reporting, i.e., that it concerns secure software development issues only after the release of the product (not with initial secure software development), was recognised above, it was also noted that external reporting can also influence initial secure software development. In theory, the reported vulnerability information enable potential buyers to make a quick and preliminary check of the security of a software product (component) they are purchasing.

However, the number of vulnerabilities reported to the public does not directly tell whether current version of the software is vulnerable or not. It only makes certain statements that hint about the issue. Especially the issue of how quickly the vulnerabilities are patched and how well does the vendor respond to the reports tells something about the seriousness of security in the development process of the specific software. When comparing software products that are both widely used, and thus presumably subject to relatively similar amount of vulnerability analysis, the number of reported vulnerabilities does tell something about the quality (and quality aspect of security) of the software[714].

---

[713] Arbaugh et al., Windows of Vulnerability, p. 57.

[714] There is a strong bias disfavouring widely used software products. The amount of vulnerability reports is likely to be much larger simply because of the wider user base. This can easily lead to the users believing that the most widely used software components are the most vulnerable ones. Even though there is no evidence of this, there is no reason to believe that the software that is in wider usage has more security related defects. This highlights the importance of understanding the nature of vulnerability reports in making comparisons between software products.

Even though historical data about the vulnerability of a software package and the responses the vendor gives to the reporters does not tell the whole story about the security of next release, it can be used as indication of the seller's attitude towards to secure software development. Thus this information can be useful in software acquisition. At least it can be useful in considering the switch to competing products. This makes the information influential for the vendor behaviour. With increased possibilities to compare the security of software products the demand for more secure software can increase.

However correct the above speculations about the capabilities and likelihoods of the users of a software product to familiarize with and to understand all this vulnerability information are, combining the limited use of the provided vulnerability information with the expenses of searching for alternatives in the market, leads to the conclusion that the benefits the increased vulnerability information provides for the competition in the software product market are likely to be low. Disclosure of vulnerability information may reduce the costs of acquiring information, but the substantial transaction costs for most customers to assimilate and use this information remain. Thus, the impact on the operation of markets may be insignificant.

This pessimistic conclusion can be resisted with the argument that if a certain informed proportion of customers use the information, then the benefits of competition will be extended to all users because vendors will not wish to lose the informed segment of its customer base[715]. In other words, it is not necessary that all users utilise the information to compare the security of different software products. Not all even have to understand or reach the information, because at the margin it is sufficient that a large enough group of influential users do[716]. Suppliers will respond if at the margin a significant

---

[715] This argument is made by Collins in *Regulating Contracts*, p. 284 and 292, with reference to Schwartz A and Wilde L (1979) Intervening in Markets on the Basis of Imperfect Information: A Legal and Economic Analysis, *University of Pennsylvania Law Review*, Vol. 127, p. 630.

[716] Of the general argument, see Ogus, *Regulation*, p. 123.

proportion of all purchasers engage in comparison shopping. When, for example, a group of big companies engage in comparing software components in terms of their vulnerability, this provides a signal to the sellers to change their products or lower their prices.

Currently only large enterprises can be assumed to be able to use the general existing scattered vulnerability knowledge. If not the vulnerability issue is separately raised in the actual purchase situation or provided in a more easily accessible form, the high search costs might hamper the more widespread use of this information. In this, the mediators like various CSIRT teams are especially helpful since they gather information about vulnerabilities on single sources that are relatively easy to access.

The already discussed lack of common metrics (scoring methods) and definitions of vulnerabilities also restricts the possibilities to compare software products on the basis of their security. When even the definitions differ and scoring varies, vulnerability information is difficult to utilise to compare different products. In addition to definitions and metrics, the comparison of software products on the basis of their vulnerability is also hampered by vulnerability information being scattered and provided in non-easily-accessible form.

In addition to the limited ability of reported vulnerabilities to enable comparison shopping, public knowledge of security related vulnerabilities can also be expected increase the likelihood of customers (including potential ones) starting to demand more secure software in general[717]. Calls for customers to rise to barricades have already been made in the sense that they should participate and campaign in favour of more secure software in different contexts. With the current high media interest the information on security related vulnerabilities increase the overall awareness of the public of the vulnerability issue and the quality and security of software; i.e.,

---

[717] At least the users can be expected to change their basic starting point in relation to the security of software products from lack of interest into assumption of vulnerabilities existing.

of the real state of secure software development in general and of the security risks involved with the use of COTS software especial.

By increasing the awareness of the lack of security in software products, the disclosure of vulnerability information can also influence initial secure software development. The users are empowered to participate in the societal efforts in achieving more secure software (such as standard setting) and to target social pressure, or to participate in civil action or in public interest organisations. The increased awareness can help in the mobilisation of social support and social pressure more widely.

# 5 Using prescriptive rules: software product liability

As it seems, the market practice in COTS software business is 'release-early-and-patch-later'. Security is added after the release of the product. Vulnerabilities that could pose a risk to security are addressed only when a breach or an external report occurs (the "penetrate-and-patch" approach) and the security requirements (features) are added as an afterthought to the standard (functional) requirements. In addition, everyone is thought to understand and expect that complex modern software will contain bugs and that this just is the way the world is.

In an ideal world of perfectly competitive market, fully informed buyers would purchase only products that provide an efficient level of quality and security. However, the substantial impediments discussed above, such as systematic misperception of accident risks, externalities and informational problems, prevent efficient operation and market outcomes are not optimal. The argument made in this study is that this industry practice for COTS software leaves significant room for improvement[719].

There are a variety of prescribed rules that are used to ease the problems in markets with physical goods (e.g., food, toys, etc.) and could be used to tackle the problems in software development that prevents it from achieving a more secure state and to minimise the

[719] Customary practice in an industry is often persuasive evidence of what is considered to be the reasonable standard of care. Conformity with the custom usually suggests that sufficient effort to address security is given. However, as has been argued above, the current practice is not optimal and, to the degree improvements are cost-justified, current practice might not reflect a reasonable level of care in software development. This is pointed out also by Chandler, *Improving Software Security*, p. 21-22, in relation to tort liability for unreasonably insecure mass-market software in the U.S.

risks posed by the security related vulnerabilities. Liability law is one of the most widely called for remedies to tackle with the problems with security related vulnerabilities in software[720]. As a regulatory instrument it crosses several legal fields. In addition to tort law it covers also areas from contract, administrative and criminal law.

The application of different liability provisions depend on the party whose liabilities are considered. The analysis stems from the potential liable parties because they are in different positions to prevent harms and consequently subject to varying requirements for the level of care required from them. In a situation where a software vulnerability has been misused to attack a system, the potential liable parties are the perpetrators who have misused the vulnerabilities, system or network administrator (including also service providers) who control either the insecure systems or the network, vendors of the vulnerable software components, or the publishers of the vulnerabilities (reporters that make vulnerabilities available to the wider public).

*Criminal liability of those misusing the vulnerabilities* to attack systems is an obvious starting point in liability law and has been taken in use by most countries[721]. However, this is only a preliminary strategy that

---

[720] A call for accountability through liability in the wide sense has been backed especially by the U.S. National Academy of Sciences. In an assembly of existing reports on computer and communications security from 2002 the Computer Science and Telecommunications Board of the National Research Council of the U.S. proposed, *Cyber-security Today and Tomorrow*, p. 15, that "[p]olicy makers should consider legislative responses to the failure of existing incentives to cause the market to respond adequately to the security challenge. Possible options include steps that would increase the exposure of software and system vendors and system operators to liability for system breaches and mandated reporting of security breaches that could threaten critical societal functions". This is by no means a new phenomenon; Peter Seipel already in 1977, *Computing Law*, p. 86, referred to a discussion on liability for "computer errors" and pointed out the differing opinions about both negligence and strict liability under tort law.

[721] According to an OECD survey from 2004 on the implementation of its widely accepted and appreciated 2002 OECD Guidelines for the Security of Information Systems and Networks, *Summary of responses to the survey on the implementation of the OECD guidelines for the security of information systems and*

targets the intentional misuses of vulnerabilities. As recognised in information security research, also unintentional acts and involuntary accidents can threaten information security without any presence of intentionality.

Criminal law has also proven to be inherently insufficient to shield the information systems and their users from attack. In an environment where anonymity and covering ones tracks is easy and the costs of searching, apprehending and prosecuting are high, the enforcement of the criminal laws has proven to be difficult. The differences in the regulatory standards and the varying practices in the pre-trial phase of the criminal process in relation to electronic evidence does not ease the situation. Coupling criminal liability with the liability for damages of the harms caused does not provide much relief since the perpetrators are often insolvent even when they are captured.

In addition, liability of the crackers only reacts to the symptoms of the underlying problem, i.e., the insecurity of software. It does not provide cure for the original problem. It can even be speculated to have negative consequences for secure software development. Extensive reliance on criminalisation can lull the users into believing that the perpetrators are deterred and no attacks occur. This false sense of security can mean that no preventive measures are taken by the users of the software. However, due to the widely recognised problems in apprehension and prosecution of computer criminals combined with the increasing awareness of insecurity of the electronic environment, such false sense of security is unlikely.

Due to the extensive research on criminal law and its wide application, there is no need to go deeper into the analysis of its capabilities to influence secure software development in this study. I am not suggesting that further work on the criminal law arena is in vain. The ongoing work on the harmonisation of the substantive and procedural computer-crime related criminal laws, like in the case of the Council of Europe Cybercrime Convention, together with the

---

*networks*, p. 5, most of the 21 OECD member countries that responded had enacted a comprehensive set of measures to combat cybercrime.

enhancement of information provision in the form of Europol and national CERTs are important. But criminal law still remains only as a bedrock strategy.

Parties that have increasingly been established with higher liability are the *users of COTS software* as components in their information systems. Current law of information security can be said to concentrate on those who are responsible of the functioning of information systems. Most of regulation on information security is found from the area of liability for information security of those processing personal data[722], governmental agencies[723] and electronic communication operators and other providers of information infrastructure[724]. They are in a position to handle much of the consequences stemming from the misuse of security related vulnerabilities. For example, they can limit the damages and minimize the exposure to risk posed by security related vulnerabilities by updating their systems and by taking protective measures like installing firewalls and virus scanners.

---

[722] Article 17 of the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ L 281, 23.11.1995, p. 31-50 (Personal Data Directive), establishes liability for the security of their information systems to those who process personal. Those processing personal data in Finland face strict liability according to the section 47 of the Finnish implementation of the Personal Data Directive (Data Protection Act 523/1999). An enhanced duty of care standard for the information security of banking services was established in the Finnish Supreme Court cases 1993:3 and 1994:80 (control liability instead of negligence and liability also for the damages of third parties).

[723] In Finland the Act on the Openness of Government Activities 621/1999 section 18 establishes a liability for the information security of governmental information processing.

[724] Article 4 of the Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications), OJ L 201, 31.7.2002, p. 37-47, set specific liabilities for the communications providers to secure their processing of personal data.

Despite them being in a position to take preventive measures against the harm occurring from software vulnerabilities, system and network administrators and other users of software components cannot directly address the causes of vulnerabilities. They can indirectly demand higher standards of quality and security from the COTS software component providers via contracting and otherwise issue pressure on COTS software vendors, e.g., by reporting discovered vulnerabilities, but the diminishing of the initial vulnerabilities in the components are typically out of their direct reach to the degree source code is not available and no rights to change it are given. And even if they would have these rights, the users of the components often lack the expertise to do this. They also face the problem of multiple version and consequent need for extensive testing of the "shirked" version when the component is updated by its vendor.

Even though criminal law liability targets also the users of COTS software components, it is private law liability under contract and tort law that primarily applies. Private law liability has especially recently been raised as one of the possible remedies for the risk posed by high technologies. Under conditions of increasing uncertainty private law liability is called for to control the risks of human decision-making in the introduction of new technologies. Reduction of uncertainty becomes one of the purposes of liability law and the control of developmental risks is central.[725]

---

[725] Western societies are highly dependent on growth based on new technologies (employment and prosperity depend on innovation). At the same time the possibilities and risks of new technologies are more and more clearly associated with human decision making which raises the responsibility of the decision-makers. However, decisions on the introduction of new technologies are increasingly made under conditions of uncertainty (cost calculations fail where decisions are taken under uncertainty: neither the extent of damage nor prevention costs can be monetarised ex ante; the damage potential itself is unknown). This makes technological innovation inevitably risky (the advantages and chances of innovation are inseparable form the inherent risks) and there is no alternative to appliers and consumers than learning by using. This argument is made by Gert Brüggemeier in The Control of Corporate Conduct and Reduction of Uncertainty by Tort Law, p. 63-65, with EU wide

A new potential liable party in the case of a software vulnerability causing harm has been found from those making the misused vulnerability public knowledge. These public *reporters of vulnerabilities* analysed in the previous chapter have been threatened with liability and have been taken to courts especially by the vendors of the vulnerable software. The argument for their liability is that increased vulnerability reporting to the public has increased the possibilities for automation of attacks and has provided means for an increased number of potential attackers to misuse the vulnerabilities.

The basis for liability is diverse. At least basic doctrines of negligence in tort law, rules against the circumvention of technological protection measures[726] and criminal law rules against viruses come

---

considerations. A leading Finnish private law scholar Thomas Wilhelmsson, Professor of Civil and Commercial Law at the Faculty of Law of the University of Helsinki who recently has been involved in the development of the European private law, in Korvausvastuu uutena sääntelyvälineenä, p. 239-265, and in *Senmodern ansvarsrätt*, p. 66-103, continues on similar lines in the Finnish legal doctrine and widens the perspective to the whole liability law instead of pure tort law.

[726] In Europe these are present at least in the Article 6 of EU Copyright in the Information Society Directive (Directive 2001/29/EC of the European Parliament and of the Council of 22 of May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society) and similar provisions in the Article 7 of the Council Directive 91/250/EEC of 14 May 1991 on the legal protection of computer programs, OJ L 122, 17.5.1991, p. 42-46 (as amended by Directive 93/98/EEC), which are not affected by the harmonised legal protection against circumvention of effective technological measures as noted in the Preamble 50 of EU Copyright in the Information Society Directive meaning that the protection of technological measures used in connection with computer programs is exclusively addressed to in Software Directive 91/250/EEC. In addition, the Conditional Access Directive (Directive 98/84/EC of the European Parliament and of the Council of 20 November 1998 on the legal protection of services based on, or consisting of, conditional access, OJ L 320. 28.11.1998, p. 54-57) also comes into question.
The legal basis in Finland is scattered and complex. At least the amended Copyright Act (404/1961) finally implementing the Copyright Directive into Finnish legislation, Penal Code of Finland (39/1989) chapter 38 section 8a, L eräiden suojauksen purkujärjestelmien kieltämisestä (1117/2001), and the renewed Act on the Protection of Privacy in Electronic Communications 516/2004, aka Lex Sonera, (same rules were also in the previous act from

into question[727]. Even though these regulations typically allow legitimate research on vulnerabilities and the publication of results, and in theory do not hamper with information security research, the vagueness of the standards together with their openness to interpretations leaves the security researches uncertain about the legal status of their work. Due to the vagueness's, these provision and especially similar provision in the U.S. Digital Millennium Copyright Act (DMCA) has been used to threaten information security research with litigation if they publish their results[728].

The dual nature of the vulnerability reports analysed in the previous chapter sets this type of liability into questionable light. Even though the reporting of vulnerabilities to the public increases possibilities for attack, the same information is crucial for the correction of the underlying bugs and for the minimization of risks. The information is also needed for the functioning of the markets for secure software. The effects of this type of liability on secure software development was considered in the previous chapter in relation to the analysis of vulnerability reporting as a regulatory instrument and is not taken further in here.

Of the different potential liable parties the concentration in here is on the currently less widely regulated *COTS software vendors*. The liability of other possible parties, like the vulnerability abusers, the users of

---

1999) come into question.

[727] For example, Council of Europe Convention on Cybercrime CETS No.: 185 and the Finnish Penal Code chapter 34 section 9a on criminal computer mischief.

[728] Public Law No. 105-304, the Digital Millennium Copyright Act of 1998, signed to law in 28th October 1998, codified in section 1201 of the United States Code title 17 on Copyrights. The problems that have appeared in practice have been described especially by the Electronic Frontier Foundation (EFF) in Unintended Consequences: Five Years under the DMCA, v. 3, September 24, 2003, available at http://www.eff.org/IP/DMCA/?f=unintended_consequences.html [14.2.2006]. EFF is a non-profit civil liberties advocacy and legal organisation for the networked world based in the U.S.

the vulnerable software components, and the reporters of vulnerabilities is analysed only to the degree they affect the liability of the software vendor (criminal liability of the vulnerability abuser as an intervening factor, potential liability of the user etc.).

Liability of the software vendors can base on several grounds. In this study, not all of the possible liabilities of software vendors are analysed. The analysis concentrates on the liability of the vendor towards customers. It is mainly about product liability where consumer protection laws like those deriving from the EC Product Liability Directive (85/374/EEC)[729] combine with both liability for damages under contract law or tort law (delictual liability)[730]. Not even liabilities of the vendors towards third parties under other doctrines of tort law beside product liability are further dealt with. Not to even speak about the possible criminal liability of the vendors, the natural persons behind them or their employees.

Product liability is a peculiar legal construct. It can come into question both in contractual relations and outside them. It is mainly about non-contractual liability under tort law and specific product liability rules such as those implementing the EC Product Liability Directive (85/374/EEC). At the same time, it can also come into question under the rules of liability for damages in the case of breach of contract rules for defective products. This is so especially in commercial relations, but even though specific product liability rules have traditionally been systematised as part of delictual (non-contractual) rules, also contract law rules and general tort law rules still play a role even in consumer product liability cases.

---

[729] Council Directive 85/374/EEC of 25 July 1985 on the approximation of laws, regulations and administrative provision in the Member States concerning liability for defective products, OJ L 210, 7.8.1985 pp. 29-33, amended by directive 1999/34/EC OJ L 141, 4.6.1999, pp. 20-21

[730] It is "mainly" about product liability, i.e., liability for damages to persons or other property than the defective product itself resulting from the use or possession of defective (insecure) products, since also damages to the defective product itself, which are covered by the general contract law rules for defective products, are considered. Actually my use of the term product liability covers damages to all property, including the damages to the defective product itself.

The concurrence, i.e., the right of the plaintiff to choose the legal category which will give the most advantageous outcome, of both contractual and non-contractual (delictual) basis for liability in the case of vulnerable software is of importance in the software context. Even though in the product liability cases involving more traditional goods than software (e.g., toys) there typically is no contractual relation between the manufacturer and the user[731], in the case of software such direct contractual relation between the software house that developed the software in question and its user is at least sough for by the use of software licenses. Thus the possible legal avenues for the user of defective software are open to choose in theory[732].

When compensation for product related damages is sought in non-contractual (delictual) relations the focus is mainly on negligence on the part of the injurer when the damage has occurred to the property of an enterprise. Breach of the duties of care established in the legal system can constitute negligence and lead to liability for damages resulting also from defective products. In non-contractual relations the focus is on the standards of care established for the producers, the security requirements of the products and the liability for damages from the breach of these requirements.

Strict liability considerations especially in the case of damages to persons or property intended for private consumption[733] have been given a central role in the liability for product related damages.

---

[731] The goods can be delivered through a long line of intermediaries like the final seller, wholesale dealer and importer, of which the user typically has a direct contractual relation only towards the seller.

[732] As we will see later on, the use of these alternative legal avenues are relatively limited in practice due to several reason. The purpose of this analysis is partly to see the real effectiveness of these possible bases for liability in influencing secure software development. In this, the factors that hamper the use of the alternative legal avenues are central.

[733] According to article 9 of the EC Product Liability Directive (85/374/ EEC) only damages to "any item of property other than the defective product itself which is ordinarily intended for private use or consumption and is used by the injured person mainly for his own private use or consumption" are compensated.

However, since the consumer product liability rules based on the EC Directive (85/374/EEC) are limited only to death, personal injuries, or damages to property (other than the defective product itself) that is intended and mainly used for private use or consumption[734], damages to other (non-consumer use) property and pure economic losses that have no relation to damages to person or property are not covered[735]. This means that in other cases than personal injuries, where special strict product liability rules apply even to the liability between enterprises[736], liability of an enterpriser against another enterprise in a non-contractual relation essentially relies on general tort law rules.

When a defect in a product has caused damage, the injured party has also the right the demand compensation from the party from which it bought the product. In contractual relations product liability concentrates on the defectiveness of products, i.e., their conformity to the contract. Breach of contract is also a requirement in the case of product liability under the default defectiveness rules in contract law. When a product is considered defective it constitutes a breach

---

[734] See article 9 of the Council Directive 85/374/EEC of 25 July 1985 on the approximation of laws, regulations and administrative provision in the Member States concerning liability for defective products, OJ L 210, 7.8.1985 pp. 29-33, amended by directive 1999/34/EC OJ L 141, 4.6.1999, pp. 20-21. As Mathias Reimann in Product Liability in a Global Context, p. 137-138, points out in an analysis of product liability rules worldwide in 2003, the U.S. approach is different from this European approach widely followed by other countries even outside Europe. In the U.S. and also in the Japan all property (other than the defective product itself), including commercial property, is covered.

[735] The importance of this in relation to software has been raised, e.g., by Stuurman in Product Liability for Software in Europe, p. 144, and in The EC Directive on Product Liability and its Application to Information Technology, p. 202.

[736] For example, in Finland all the damages to persons caused by defective products are covered by the Finnish Product Liability Act (694/1990). Thus also personal injuries of the enterpriser and her employees are covered by the Product Liability Act. However, as Doctor of Laws Marko Mononen who currently is associate in a private law firm and Docent of Civil Law at the University of Helsinki, notes in 2004, *Yritysten välinen tuotevastuu*, p. 134, the injured person has a freedom to choose the most favourable legal basis for her claim.

of contract and can lead, among other remedies like repair of defective good, reduction of price, withdrawal of a contract, which are primary in most of breach of contract cases, also to liability to pay damages.

When compensation for product related damages is sought under contract law, it is the contract that forms the primary legal basis. Liability for defective products, especially when considered in a relationship between enterprises, is considered on the basis of the contract between the parties. In the case of damages to persons or property (other than the defective product itself) ordinarily intended for private consumption, the indispositive rules in the Product Liability Directive (85/374/EEC) and implementing legislation together with the national consumer contract law rules, that cannot be agreed in the disfavour of the consumer (i.e., they are indispositive), provide the primary legal source.

Even though software vendors currently are not subject to much liability regulation, attention on the issue has increased and demands for software vendor liability has been made on several occasions in recent years. Especially practitioners of computer security have been active in calling for increased liability[737]. Software vendor liability has not only been called for in the U.S where the role of private law liability in general and product liability especially is higher than in Europe in general due to the different social insurance systems. Also in Europe, for example, the Economic and Social Committee of the European Communities, a European level forum for reflection of civil society organisations and associations, in its opinion on the European Commission Communication on Network and Information Security stated that "[b]usinesses and private individuals must therefore be given more effective legal means to make software

[737] Bruce Schneier has been one of the most active advocates. Of the many occasions where he arguments in favour of legal liability for software vendors see, Schneier B (2002) Liability and Security, Grypto-Gram Newsletter, April 15, 2002, at http://www.schneier.com/crypto-gram-0204.html#6 [23.2. 2006]. See also Declan McCullagh, *A Legal Fix for Software Flaws,* CNET News.com, August 26, 2003, at http://news.com.com/2100-1002_3-5067873. html [23.2.2006] (quoting computer experts calling for software publisher liability and legal experts who are more doubtful).

operators and manufacturers financially liable for serious security and data protection lapses ascribable to them under product liability legislation"[738].

It must be stressed that the analysis of software vendor liability concentrates on the function and structure of liability law as an incentive mechanism to promote product safety and security. This is *not* a search for certain interpretations of the valid (positive) law in the sense it is done in prescriptive legal scholarship as such. The details of the various rules are not analysed and the statements apply both to EU and Member State levels, even though the emphasis is on the Nordic perspective due to the author's background in Finnish legal system[739]. Instead, the focus is on analysing the capabilities of liability rules to influence secure software development.

At best, this is an analysis of the interpretative space of valid product liability rules in order to see if the positive law could be used to enhance secure software development. The normative argument that product liability provision ought to be either interpreted or amended in a way that promotes the security of software is just one practical argument among several competing and possibly even stronger arguments for the interpretation of the rules and for their amendment.

---

[738] Opinion of the Economic and Social Committee on the Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions on Network and Information Security: Proposal for a European Policy Approach, COM(2001) 298 final, TEN/083, Brussels, November 28, 2001, section 3.1.11, available at the home pages of the EESC at http:// www.esc.eu.int/ index_en.asp [23.2.2006]

[739] The analysis of the law will concentrate on EU level law and take examples from one member state (where deemed appropriate). The EU law seeks to harmonise the differences between national laws of Member States. Even though national law and EU law are mutually dependent, EU law takes precedence over national law.

## 5.1 Who regulates ?

In the case of product liability based on special laws or general contract and tort law rules the regulators are diversified. There are state-centric regulators using legislation and case-law, legal scholarship constructing systematic arguments, market actors constructing contractual practices for the specific line of business and the contracting parties as such altering and choosing the laws applicable to their relation[740].

The strongest form in which the state actors regulate is when mandatory (indispositive) rules that disregard or change the terms of contracts are used. This is the case typically when a weaker party, such as a consumer or a SME, needs protection. A central example is the consumer protective product liability regulation, where contracts for less security than the regulation allows are deemed invalid. However, state-actors can be seen as regulators even when only dispositive liability rules are used.

Even though dispositive private law rules especially in the area of contract law are conventionally seen as having mainly a facilitative role, i.e., they permit citizens to make legally enforceable contracts and thus lowers transaction costs[741], the purpose of private law is

---

[740] Specifying the regulators even at the level of state actors is tricky since much of product liability regulation outside consumer protection in Europe is a construct of legal scholarship and case law rather than of specific state-centred regulatory action.

[741] According to the conventional view, the law of contracts facilitates market transactions as, e.g., Professor of English Law at the London School of Economics and Political Science Hugh Collins points out in Regulating Contract Law, p. 13, in which he continues on the inspiring analyses of contract law combining fine contractual scholarship with empirical studies and social theory, thus forming a model of the type of scholarship that combines regulatory and doctrinal research. According to the economic explanation, as made explicit in basic textbook on law and economics like that of Robert Cooter and Thomas Ulen, *Law and Economics*, p. 199-202, the dispositive general contract law rules function in the background as a collection of the rules of liability that are applied when the contracts or the contractual practices in the line of business are silent or otherwise give no guidance on the issue. They provide a default allocation of risks between the contracting parties that

increasingly being seen to control behaviour by reference to enforceable standards in the pursuit of a particular goal.

This is relatively widely acknowledged in relation to contract law[742]. Considerable controversy exists over the goals that tort liability should serve and doctrines in tort law do not seem to have a definitive stated purpose, which hinders the analysis of tort law as a regulatory mechanism. However, its regulatory role has been widely accepted[743]..

At the same time also the contracting parties are acting as regulators[744]. The private actors can determine through their agreement what legal regulation, if any, will apply to their transaction[745]. To the

---

fill gaps left to the contracts. This reduces transaction costs in situations where the default terms provided by law are preferred by both parties, since they can omit these terms from the contract.

[742] The regulatory function of contract law has been argued especialyl by Hugh Collins in Regulating Contract Law, p. 17, and by Jan Hellner in *Lagstiftning inom förmögenhetsrätten*, p. 162-164.

[743] This has been pointed out by Jane Stapleton, Professor of Law in the Research School of Social Sciences at the Australian National University and Ernest E. Smith Professor of Law at the University of Texas School of Law, in a critical evaluation of the susceptibility of tort law to regulatory analysis, Regulating Torts, p. 130-134. See also Schuck, Tort Liability, p. 466. Even those tort scholars that reject the deterrence arguments as valid explanations of tort doctrines and favour corrective justice arguments (i.e., reject the argument that deterrence is a valid purpose of tort law), accept that tort law can influence behaviour and can be used to do so as a recognised U.S. tort law scholar Gary T. Schwartz argues in Mixed Theories of Tort Law, p. 1827, an article where he reconciles the doctrinal and the economics-based theories of tort liability.

[744] As radical is this may seem at the first hand for doctrinal lawyers outside contract law and for those who perceive regulation as a state-centric activity, there actually is nothing new in this statement. Legal theory has for long recognised that also private individuals regulate and that even ordinary people can produce legal norms. For example, HLA Hart in a classic of legal philosophy and jurisprudence, *The Concept of Law*, p. 41, holds that human agents have the capacity to become private legislators.

[745] This ability is based on the private power-conferring (constitutive) rules; secondary rules of change in the Hartian conception, that confer power on individuals to vary their initial legal positions (Hart, *The Concept of Law*, p. 95-96). These power conferring rules also regulate; inside the behaviour they have constituted they also guide behaviour. Even though they are rules of the game, e.g., how to make a contract, they also regulate behaviour by showing how the

extent that legal regulation is comprised of default or supplementary (dispositive) rules, the parties are permitted through express terms to prefer their own standards over those supplied by the dispositive legal norms[746]. As argued by Hugh Collins, it is possible for the contracting parties to structure their relationship in ways that avoid the application of such traditional regulation of contractual terms even when mandatory rules are used[747]. Thus the regulatory capability of contracting parties is wide.

Software business in general and COTS business especial widely use their own contract templates to facilitate the contracting procedure and to reduce transaction costs[748]. The drafters of these standard templates are central regulators. They can be drafted unilaterally by one of the contracting parties or by an industrial organisation representing only one the parties. A typical example is a seller or an organisation representing the sellers. Bilaterally or multilaterally drafted contract templates, agreed documents, are the product of the representatives of both of the contracting parties.[749]

---

society wishes the game to be played.

[746] Hugh Collins, Regulating Contract Law, p. 16, develops this argument in relation to contract law into a form where the private law of contracts in using dispositive norms, and respecting the private autonomy of the parties, delegates to the parties the power to fix the rules that will govern the relations between the participants in the market.

[747] Collins, Regulating Contract Law, p. 16-17.

[748] As noted by Juhani Warsta in an analysis of contract processes and relationships in software business from 2001, *Contracting in Software Business*, p. 53, standard contracts in software business are used because it is considered to have its own characteristics that require special rules to address them. They are used to replace the default rules provided by law that serve as background rules to which the contracting parties can rely on and leave matters open, and thus save transaction cots.

[749] Thomas Wilhelmsson in *Vakiosopimus*, p. 31-32, applies this argument generally to all standard form contracts. Agne Lindberg and Daniel Westman in *Praktisk IT-rätt*, p. 278, show that both uni- and multilaterally drafted standard terms are used also in IT-contracts in Sweden.

## 5.2 Influence mechanism

The mechanisms by which liability rules, especially product liability, influences secure software development and behaviour in general are manifold and highly speculative. Certain basic conceptualisation can, however, be made. Of the basic mechanism by which regulation influences behaviour, both external constraining and intrinsic predispositions shaping can be identified.

*Internal influence.* Liability rules for defective and insecure products, i.e., product liability rules understood widely to include doctrines also from contract and tort law in addition to the consumer oriented strict product liability laws, are authoritative norms that provide general standards of conduct that communicate of the desired behaviour for the objects of regulation[750]. They communicate in general concepts, which are specified in concrete liability court cases or contracts, what kind of behaviour and level of activity is expected from the objects of regulation. Especially the clarification, restatement or even the promulgation of a totally new standard for conduct in a product

[750] Basically, product liability rules prescribe a body of substantive legal doctrine (i.e., rules and principles) that define the legally enforceable duties that an individual or an entity owes to others and the circumstances in which she may violate those duties (i.e., the standard of care). Note that strict liability differs in the sense that no standard of conduct is provided. Liability results from all activity that causes harm despite the level of care taken by the injurer. Liability is based on damaging conduct. However, product liability regulation in the consumer protection side (e.g., EC product liability directive 85/374/EEC) is not strict in this sense. Liability is qualified in the sense that it does not cover all the damages caused by the product. Liability is non-fault based, i.e., not focused on the conduct of the producer, but it is tied to the product characteristics. Liability arises from the damage caused by a defect in the product, i.e., when it does not provide the safety that a person is entitled to expect. This brings the standard of conduct into the European product liability regulation.

liability lawsuit[751] and the subsequent publicity given to it could induce software vendors to re-examine their development practices.

These standards for conduct as such, without sanctions necessarily being attached to them, can prevent defective products from entering markets and causing harm. This is the expressive influence where information on liability rules influences intrinsic predispositions via the change of social norms or social forms. Promulgation of a law or a new rule established in a court case give people instructions on what to do (i.e., the meaning of the norm, the behaviour it expects from the addressees, is expressed), their justification enhances legitimisation and by means of convincing enhances internalisation.

In the legal analysis of tort law this type of prevention is especially attached to the moral building or up-keeping of social or moral norms inside the class of regulatory objects. New promulgated liability provisions that set standards for conduct affirm the blamefulness of the regulated behaviour. They can also represent, when made according to democratic principles, society's moral norms, and, in the case of rules created via litigation, serve to alter the behaviour of also other parties[752] than those represented in the litigated case

---

[751] Recent argumentation in liability law also in the Nordic doctrine, especially by Thomas Wilhelmsson in *Senmodern ansvarsrätt*, increasingly recognises the possibility of constructing new standards (norms) via litigation as a regulatory mechanism. Preventive effect of liability litigation is not related solely to the breach of rules explicitly set elsewhere, but litigation can also create altogether new rules (Wilhelmsson, *Senmodern ansvarsrätt*, p. 53). However, as Wilhelmsson in *Senmodern ansvarsrätt*, p. 186, points out, liability litigation used as a tool for micro politics in this new rule generating form involves numerically marginal cases since majority of tort litigations involve small damage cases with more common issues. It must be noted that in this case it is not the litigators, be they private individuals, NGOs or others having standing in court, that create the new rules. Instead, it is the court that formulates the new rule and acts as a regulator.

[752] The learning process that occurs between the parties in a court case does not necessarily involve a wider moral discourse. Litigation, at the latest, forces also the injurer to consider its behaviour in comparison with the legal rules and, if the case is not settled, the court then restates the standard and makes explicit the related costs. This can lead to changes in future behaviour in the form of higher precaution or even lead to the internalisation of the rules as

especially when the rules are created in high media profile court cases[753].

Suggestive empirical evidence of such expressive effect can be found from the considerations of the various parties in product liability claims. For example, in an EU study on product liability from 2003 John Meltzer, Rod Freeman and Siobhan Thompson report that the most important factors behind the increase in the number of consumer product liability claims are considered to be increased consumer awareness of rights, increased consumer access to information, and media activity[754]. They all point to the same direction; increase in the information about the possibilities for consumers to receive compensation for product related accidents. Whether this is through increased information provision *on* product liability regulation (especially information on Product Liability Directive and implementing national laws) by the regulator or higher media attention in product liability cases was not explicitly examined in the EU

---

part of the intrinsic predispositions of the objects of regulation. This is special prevention in the sense as it is used in criminal law but its influence mechanism is the same as standard economic prevention. This usage of the terms is different from the occasional identification of economic prevention with special prevention at least in the Finnish doctrinal tort law analyses as noted by Antti Kivivuori in Vahingonkorvausvastuun tarkoitusperät, p. 170.

[753] The learning process in the form of moral education through litigation, depicted in the Finnish doctrine especially by Thomas Wilhelmsson in *Senmodern ansvarsrätt*, p. 103-134, can also happen without much publicity through the system of precedents or through information provision to the legislators and other regulators (idem., p. 161-162). In addition to these influence mechanisms, the ability of tort law to affirm and reinforce society's essential norms is also explained especially in the common law countries by its reliance on the jury system, through which tort law can track society's basic beliefs about right and wrong conduct (Schuck, Tort Liability, p. 477-478).

[754] Meltzer et al., *Product Liability in the European Union*, p. 32-34. The respondents to the survey consisted of consumer representatives (i.e., consumer associations both at national and EU levels), producers and suppliers of products together with trade associations, insurers and their associations, and representatives of the legal field (both claimants' and defendants' lawyers, regulators and other governmental agencies, and legal scholars) from all Member States (Meltzer et al., *Product Liability in the European Union*, p. 6-7).

study[755], but both of them belong to the expressive influence mechanism used in this study.[756]

*External constraint.* Various sanctions (remedies for the injured party) are typically attached to the product liability rules either under tort and contract law. These remedies, such as monetary damages, and specific performances like repair of defective good, reduction of price, withdrawal of a contract, which are primary in many breach of contract cases, function as additional incentives for the producers in achieving the standard of conduct for the development of high quality and secure products prescribed in the liability rule. They have an additional preventive effect on harms caused by defective (insecure) products. This is the external constraining provided by liability rules[757].

---

[755] Due to the difficulties in the formulation of empirically testable hypothesis about the expressive influence, it is not surprising that no further research on these issues were made. Meltzer, Freeman, and Thompson do, however, in their study *Product Liability in the European Union*, p. 32, recognise the possibility that the implementation of the Directive may have contributed to an increased consumer awareness of rights, and that media activity would be expected to play a role in providing consumers with access to information.

[756] Even though this necessarily does not directly apply to the secure software development context due to the several contradictory messages on the applicability of product liability rules on software defects in general and security related defects especial, which will be discussed later on, it shows the existence and possibilities of such influence mechanism.

[757] Note that the costs do not necessarily result from the enforcement of legal sanctions. As Hugh Collins in *Regulating Contracts*, p. 123-5, explains in relation to the role of the adjudication process in the construction of markets and in the regulation of contracts, regulation through adjudication can influence behaviour even when no sanctions are provided, when the court serves the function of collecting and disseminating what participants in the market regard as reliable information about trustworthiness. This enables the use of non-legal sanctions such as damage to reputation. However, because these costs can incur over businesses even without actual legal liability, they cannot be considered as part of the costs that belong to the influence mechanism of the liability rules as such. In trying to clean out the influence that the law as such has on behaviour (the normative restriction made in the analysis of the effects of laws), considerations of several costs related to the marketing of defective products need to be bypassed. They are not attributable to the liability rules as such in the form the analysis of normative guidance mechanism requires. They

Note that the users of the goods can typically also take measures to prevent accidents; this is called bilateral or joint precaution. Especially so in cases involving misuse of security related bugs in software, e.g., by taking protective measures like firewalls and virus-scanners and keeping their software updated. Also their behaviour depends on the level of liability. However, the regulatory analysis concentrates on the vendors of the software who are the main objects of regulation in the case of product liability rules. The possibility for bilateral precaution is reflected in the defences against product liability present in the regulation.[758]

Whereas the former type of prevention is attributable to the mere content of the rules, their promulgation and justification, together with other forms of information provision about the standards of conduct, this latter form of prevention is typically modelled in

---

can be reasons for action even without the existence of the liability rules. They will be, however, picked up later on in the analysis of the factors shaping the influence of product liability rules.

[758] Because the probability for private individuals to face significant losses due to security related defects in software is relatively low, they might mistakenly assume that the probability of an accident is zero (will never happen) and take very little precaution, as suggested by academic literature (Cooter and Ulen, *Law and Economics*, p. 330, with reference to Kahneman D and Tversky A (eds., 1981) *Judgement Under Uncertainty: Biases and Heuristics*). That fact, if established, could make this a situation of unilateral, rather than bilateral, precaution, i.e., only vendors and implementers could realistically be expected to take steps to reduce the probability and severity of accidents as pointed out by Robert Cooter and Thomas Ulen in *Law and Economics*, p. 331. However, the precaution can be seen as bilateral even in consumer-enterprise relations despite these possible errors of private individuals in making calculations about risks and consequent preventive actions. The probability is increasing and several widely published cases of damage due to viruses, worms, trojans and crackers make such issues seem even more likely. Note that for well-publicised, potentially catastrophic outcomes, most people systematically exaggerate the probability of an accident occurring regardless of the objective information to the contrary (Cooter and Ulen, *Law and Economics*, p. 330, with reference to Kahneman D and Tversky A (eds., 1981) *Judgement Under Uncertainty: Biases and Heuristics*).

economic terms and is attributable to the additional costs incurred by the sanctions attached to the breach of the rules[759].

Modelling of behaviour done especially in law-and-economics suggests prevention in the economic form where the threat of civil litigation forces the producer to consider the expected social costs caused by defective products[760]. When potential producers of defective goods internalise the costs of the harm they cause in this way, they have incentives to invest in preventive measures at the socially efficient level. The possible costs incurred by liability in the case of defective products causing harm function as incentives for the software vendors to make software that is as secure as the communicated standard requires[761]. This deters them from making excessively defective products[762].

---

[759] Naturally information is necessary also for the economic prevention side of liability rules as pointed out by Göran Skogh in *Priser, skadestånd och straff*, p. 48.

[760] It is the foresight of being held liable *ex post* that induces parties in the accident setting to take preventive measures. Actual litigation is not necessary for liability costs to influence decision-making as noted by Daughety and Reinganum in Product safety, p. 1188. Gary T. Schwartz in Mixed Theories of Tort Law, p. 1816-1817, has formulated this preventive effect in relation to tort law in the following manner: "So long as the defendant can anticipate bearing liability for the range of injuries his tortious conduct foreseeably can produce, this prospect can induce in him an appropriate deterrence response". In the COTS software context, this deterrence approach can be considered as appropriate since the regulation is directed at commercial enterprises that, at least, ought to be cost calculating organisations. They come closer to the central assumption in economic theory about the rational self-interestedness of decision-makers, which is vital for the tort liability system to be able to send signals to potential victims and injurers about the desired behaviour. The vitality of the connection between the assumption of rationality and the economic model of tort liability is discussed, e.g., by Robert Cooter and Thomas Ulen in *Law and Economics*, p. 330-331.

[761] The underlying assumption is that businesses observe the regulation as a cost which needs to be minimised by cost-effective adjustments to their operations.

[762] The assumption still is that non-defective products cannot be made. This cannot even be the goal of regulation. Instead, optimal deterrence is the objective. Defective products continue to be produced, but only to a certain acceptable threshold communicated in the liability rule.

> Note that there are, as Gary T. Schwartz explains, both economic and non-economic ways of understanding deterrence[763]. The deterrence that product liability rules provide can itself be understood not just as a maximiser of utility but also as a device for achieving justice or, at least, as a device for reducing injustice[764]. However, this does not alter the influence mechanism. Only the final objective of regulation is understood differently. While serving as an instrument of deterrence, product liability rules ultimate objective becomes the achievement of justice or, at least, the prevention of injustice, instead of maximisation of utility[765].

Empirical evidence from product liability regulation also verifies this latter form of influence. Most participants in an EU study on the practical operation of the systems of law under which product liability claims may be brought in the Member States, for example, thought that the Product Liability Directive (85/374/EEC), since its adoption in 1985, had moderately increased the prospects of product liability claims being brought, and of their success[766]. Most participants also

---

[763] Schwartz, Mixed Theories of Tort Law, p 1802 and 1831-1832.

[764] When accepting the perception of corrective justice scholars that it is unjust or unfair for defendants to inflict harm on plaintiffs through their negligent behaviour, which is used to explain why the goal of correction is achieved by allowing the plaintiff to secure compensation ex-post from the defendant (Schwartz, Mixed Theories of Tort Law, p. 1831), it becomes evident that tort law can, by requiring potential defendants to consider the prospect of liability, deter those defendants from engaging in negligent behaviour and hence bringing about an unjust result. In particular, tort liability can properly be seen as a practice designed to deter defendants from violating the moral rights of potential victims. If the goal of liability rules is to strike a balance between the vendors liberty of action and the users right to security, then, certainly, preventing accidents from happening does a better job in affording security than merely offering compensation once an injury has occurred.

[765] According to Schwartz, Mixed Theories of Tort Law, p. 1831, this deterrence-as-justice rationale does not fully apply to rules of strict liability. They cover cases in which the act of the defendant that results in injury is not necessarily improper. However, in product liability it is likely to be so, since the liability is based on the expectation of security by the users of the product (i.e., a standard of care).

[766] Meltzer et al., *Product Liability in the European Union*, p. iii and ix. Note that

thought that the Directive had contributed to an increase in the level of safety of products in the EU, even though only to a limited degree[767]. Similar findings have also been presented in empirical studies using a variety data sources on the effects of product liability for the safety of products in other countries[768].

*Duty to disclose.* The preventive measures to which the liability provisions incite to by the threat of liability (additional costs) are not restricted to the improvement of the security and quality of the product as such. Even though it is the main assumption made by the regulators, research on liability regulation and risk management show a variety of preventive measures that are taken by the objects of

---

data about the reality of product liability, i.e., about claim numbers, litigation rates, win-lose-ratios, and actual awards, are scarce and hard to come by with regard to most countries in the world. Thus, assessments of the law in action are only indicative.

[767] Even though most of those who expressed their view on whether or not the Product Liability Directive had contributed to increasing the level of safety of products marketed in the EU considered that it had done so only a little, it is noteworthy that approximately one third thought that it had not contributed at all. Majority of the producers considered that the Directive had not contributed at all (Meltzer et al., *Product Liability in the European Union*, p. 25 and 26). This is also emphasised by Mathias Reimann in Product Liability in a Global Context, p. 145, on the basis of existing studies combined with national reports received for the compilation of the General Report I represented at the XVIth Congress of the International Academy of Comparative Law in Brisbane, Australia, on 18 July 2001. The evidence about the reality of product liability, i.e., law in action, despite of its deficiency and imprecision, creates a noteworthy overall impression: in most countries in and outside Europe, the influence of the EC Directive has not really made much of a difference in practice.

[768] Mark Geistfeld in Products Liability, p. 360-362, provides an overview and analysis of several such studies. As John Meltzer, Rod Freeman and Siobhan Thompson in *Product Liability in the European Union*, p. 54, note the evidence of practical experience of the Product Liability Directive still remains limited, even after almost 20 years since it was first implemented by a Member State. Only some clear trends can be identified together with a number of general conclusions concerning the operation of the Directive.

regulation in order to manage product liability risks[769]. Without going deeper into the ways the objects of regulation can actually react to the liability risk in here, which will be discussed later in relation to the reactions of the objects of regulation, another main way in which the regulators seem to expect liability rules to influence behaviour can be identified. In addition to the above mentioned influence mechanism, product liability rules influence secure software development by setting duties to disclose security information[770].

In products liability, information provision has a special role. Since both the vendors and the users can take preventive measures against the accidents occurring due to software vulnerabilities[771], the party who posses this information is obligated to inform. It is the vendors in market-driven development, like all product manufacturers as argued in economics theory by Robert Cooter and Thomas Ulen[772], who are in control of the design of the products and their development process, and who are most likely to be aware of any special dangers their products present and, therefore, can most efficiently convey information about those dangers through warnings. The purpose of product liability rules can be seen to establish strong incentives to the producers not only to reduce their liability for accidental harm by making their products safer in normal use, but

---

[769] Mika Hemmo in Tuotevastuuriskit, p. 31-49, provides a short overview of the many possibilities by which companies can manage their product liability risks. Marko Mononen, *Yritysten välinen tuotevastuu*, p. 321-338, does the same in relation to business product liability, and Thomas Wilhelmsson and Matti Rudanko, *Tuotevastuu*, p. 265-301, in relation to consumer product liability.

[770] The fact that product liability rules together with marketing laws create duties to disclose safety and security information is the reason why vendors' information provision is treated separately from other possible risk management techniques. Information provision is explicitly incited to in the rules, even though the scale of preventive measures to which the product liability rules incite to in general is much wider.

[771] Also users can take precautions to reduce the probability and severity of accidents resulting from vulnerabilities in software, for example, by keeping their software updated, installing the latest patches, running anti-virus software and setting a firewall. This situation is called bilateral precaution.

[772] Cooter and Ulen, *Law and Economics*, p. 358.

also to inform users of the increase in the risk that would result from various forms and degrees of carelessness by the user and of the inherent risks of their products in proper use.

Both types of information are essential inputs into efficient decisions on product safety by the users as noted by Paul Burrows[773]. However, in the case of software security, information on the product's inherent susceptibility to risk when it is properly used is the more relevant type of input into efficient decisions on product safety by users. As noted by Burrows, in the case of risks attached to the improper use of products, duties to disclose play only a residual role, i.e., to provide producers with the appropriate incentive to warn for products not covered by statutory warning requirements[774]. However, since there are no statutory hazard warning requirements established for the vendors of COTS software, the general duties to disclose are especially important.

> Of course information on the increase to the risk that would result from various forms and degrees of carelessness by the user is also central, like in the case of using COTS products for high security purposes or putting input that is not intended into the program (remember the black-box testing where various forms of input is used to analyse when the product crashes). Failure of the user to head to warning or to follow the instructions provides a basis for the defence of contributory negligence for the vendor. The information provided is of a kind to enable the user to reduce the accident risk by taking care in an efficient ways.[775]

---

[773] Burrows, Products Liability and the Control of Product Risk in the European Community, p. 79.

[774] Burrows, Products Liability and the Control of Product Risk in the European Community, p. 80.

[775] As the economic analysis of the law suggests (Cooter and Ulen, *Law and Economics*, p. 358) strict liability with the defence of assumption of the risk and product misuse (defence of comparative/contributory negligence) is an efficient standard for minimizing the social costs of product-related injuries.

The disclosure requirements set by product liability rules come up at many points in such questions as whether one must inform the opposite party in a negotiation so as to avoid him contracting under the spell of an error (otherwise the contract is not enforced) or whether the seller must disclose apparent defects in his products (or assume liability when accidents occur)[776]. The duty to disclose safety information is most clearly visible in consumer product liability laws. For example, according to article 6 of the EC Product Liability Directive (85/374/EEC) a product is considered to be defective "when it does not provide the safety which a person is entitled to expect". In the consideration of the expectations the presentation of the product, including instructions for use and warnings are relevant. When sufficient instructions and warnings are provided, the product is not considered defective. Thus a duty to disclose safety information is created.

Duty to disclose safety information is present also in commercial product liability law under general contract and tort law[777]. For example, the contractual liability rules for defects in products are based on the assumption that the buyer can seek for contract law remedies for defects in products only if she did not know about the existence of such defects. Under tort law, the duty to disclose is present especially in the failure to warn doctrine; failure to warn of the

---

[776] Note that the provided information can also be a ground for discharge from liability. It affects users' knowledge of risks and the level of care required from them. Failure of the user to head to warning or to follow the instructions provides a basis for the defence of contributory negligence for the vendor. This has been given less significance in tort law as Marko Mononen, *Yritysten välinen tuotevastuu*, p. 173-174, points out in relation to the Finnish doctrine. See also the renewed textbook on Finnish product liability law of Thomas Wilhelmsson and Matti Rudanko, *Tuotevastuu,* p. 276-281.

[777] Similar disclosure requirements are set also in general marketing rules against unfair advertising or marketing both in consumer and commercial contexts. In the Finnish legal doctrine Thomas Wilhelmsson and Matti Rudanko in *Tuotevastuu*, p. 266, classify marketing regulation as a part of the preventive regulation of hazardous products together with product liability and product safety regulations.

accident risks present in released products is considered to be negligent behaviour.

In relation to secure software the disclosure requirements concern especially vulnerability information. In the case of software vulnerabilities, the legal duties to disclose are derived mainly from the same product liability regulations, general contractual and delictual (non-contractual) liability rules for defects in products that set standards for conduct in relation to security of products in general. In addition, duties to disclose are created by doctrines of pre-contractual liability (like culpa in contrahendo) and general marketing rules. Under these general rules on defects in products, negligence, and marketing regulation both in consumer and commercial relations, and especially in product liability laws, the vendor must provide safety information concerning their products or assume liability when accidents occur. The informing of the defectiveness of products and the risks involved is for the vendor a way to minimize or even avoid liability.

The rules that allocate information in general stem mainly from regulatory law. They have an especially central role in consumer protection laws. In relation to safety and security information, the rules stem especially from the statutory hazard-warning requirements under general and sectoral product safety regulations such as Directive 2001/95/EC on General Product Safety[778]. However, there are no

[778] Directive 2001/95/EC of the European Parliament and of the Council of 2 December 2001 on general product safety, OJ L 011, 15.1.2002, p. 4-17. These informative instruments need to be separated from the duties to disclose in contract, tort and specific product liability laws. The division of regulatory tools that stimulate information supply into duties of disclosure and statutory hazard-warnings is presented by Paul Burrows in Products Liability and the Control of Product Risk in the European Community, p. 79. The difference is that the statutory warning requirements established in general and sectoral product safety regulations are backed up by an administrative control system and the threat of criminal law sanctions (fines) in the final-end. These regulatory instruments are not taken further in this study. They essentially belong to the informative category and play a marginal role in relation to software security. The administrative control system of product safety regulations is currently not capable of dealing with security issues in COTS software. And is not likely to do so due to the marginal role COTS software

clearly stated statutory hazard-warning requirements for software vendors, i.e., mandatory disclosure rules of standardised information, in relation to software vulnerabilities[779]. This is why the general liability rules for defective products both in consumer and commercial contracts, and especially product liability regulations, together with general marketing rules are central. This is also the approach taken in the first Finnish study on information security and law from 1997, where the normative argument was made that the trader of digital products have on obligation to inform about the information security risks related both to the operational environment and the marketed product, even without the support of specific disclosure rules[780].

In Finland, the obligation to know the marketed product and the duty to inform of the risks of which the seller knows or should have

---

plays in the physical safety of consumers and the general lack of knowledge in software security issues that hampers the possibilities of the administrative control system to handle them.

[779] Note that especially in electronic communications the system operators have obligations to inform their subscribers of particular risks of a breach of a security of a network under the article 4 and especially subsection 2 of the Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) OJ L 201 , 31/07/2002, p. 37-47. See also the preamble 20 of the Directive. This duty to inform concerns only risks in the systems of the service providers and thus does not apply to software vendors as such (assuming they do not provide such electronic communication services). However, since the particular risks to be disclosed can be due to the vulnerabilities in the components of the systems, such as a standard off-the-self firewall, the service providers have a limited obligation to inform also of such risks. This not only makes the existence of the vulnerability in the system widely known (negative information) but also increases the costs of managing the disclosure.

[780] Saarenpää and Pöysti, *Tietoturvallisuus ja laki* [Data Security and Law], p. 462-463. This is also the approach of the Finnish Consumer Agency. They advice the providers of broadband connections to inform their customers of the need for security measures such as firewalls, anti-virus software and updated operating systems in the open networks. Of the previous recommendation see the Press Release 24.11.2004 of the Finnish Consumer Agency at http://www.kuluttajavirasto.fi [23.2.2006]

known do not have to based on specific legislation. However, the general rule developed in legal practice has been expressed in addition to the product liability rules also in special consumer marketing provision[781].

The consumer marketing provision in the Finnish Consumer Protection Act (38/1978) chapter 2 section 1 subsection 2 indirectly obligates traders to convey information necessary in respect of the health or economic security of consumers. However, also the general clauses on unfair marketing in the first subsection of the above mentioned rule and similar rule in the Finnish Unfair Business Practices Act 1978 section 1 have relevance also in relation to disclosure of safety information; keeping a vulnerability secret might constitute unfair marketing or business practice. They provide an indirect obligation because it based on a conditional injunction; the trader is forbidden to market a good unless she gives such necessary information. These provisions are from time before specific consumer safety or product liability regulations; they have now been re-stipulated in them as pointed out by Thomas Wilhelmsson[782] and Mika Hemmo[783].

In the case of duties to disclose under contract law it is important to differentiate between the types of information required to be disclosed[784]. Market exchanges are often motivated by differences

---

[781] This has been explicitly stated in the Finnish Supreme Court tobacco case 2001:58.

[782] Wilhelmsson, *Suomen kuluttajansuojajärjestelmä*, p. 126-128.

[783] Hemmo, Kuluttajamainonnan informatiivisuusvaatimuksista, p. 368-371.

[784] This grouping of types of information to be disclosed presented in the economics of contract law (see, Cooter and Ulen, *Law and Economics*, p. 273-275) is also illustrative in relation to tort law. Only relevant information category in relation to the duties to disclose for tort law is safety information. Note that this type of substantive approach to the disclosure problem where the permissibility of nondisclosure is decided on the basis of the character of the information is by no means to only one. As Melvin A. Eisenberg points out in Disclosure in Contract Law, p. 1664, the stage setting work on the disclosure problem, Anthony T. Kronman's Mistake, Disclosure, Information, and the Law of Contracts, in the Journal of Legal Studies, 7(1): 1-35, 1978, utilised a process approach where the process through which the information

in the contracting parties' private information[785]. As a general starting point the gathering of *productive information*, that can be used to produce more wealth, is incited by allowing its free utilisation. An example of productive information could be a discovery of new way to avoid vulnerabilities in software development. The treatment of *redistributive information*, that creates a bargaining advantage that can be used to redistribute wealth in favour of the informed party, is different. As a general rule, the deliberate discovering of such information is discouraged. A typical example of the latter is insider information in stock trading. Even though most information is both productive and redistributive, contracts based on difference in such mixed information are also typically enforced and the investment in discovering such information is thus rewarded. Under the liberalistic standard contract doctrine in EU, which somewhat differs from the more socially oriented Nordic doctrines in that it is has a more market-rational basic approach as argued by Thomas Wilhelmsson[786], the obligations to inform are exceptions and there is no general duty to disclose information to the uninformed party[787]. However, *safety*

---

has been acquired is decisive: if the information has been acquired by a deliberative process, disclosure is not required; if the information has been acquired by a casual or an adventitious process, disclosure is required. The process approach works for casual or adventitious acquisition, but as Eisenberg in Disclosure in Contract Law, p. 1663, it is too broad when it comes to deliberative acquisition. The substantive approach becomes relevant in the cases of deliberative acquisition. For a deeper discussion of the economic discussion on the substantive approach, see Eisenberg, Disclosure in Contract Law, heading III.

[785] As Cooter and Ulen, *Law and Economics*, p. 271, in their basic textbook on law and economics argue, this difference is often solved by private bargaining, e.g., by offering to buy the information or the resource in the use of which the information is needed.

[786] Wilhelmsson, *Social Contract Law and European Integration*, p. 211.

[787] The Nordic contract law doctrine under fair dealing and good faith in contract has been moving from the free utilization of information towards wider obligations to inform the contracting party as pointed out by Mika Hemmo *Sopimusoikeus I*, p. 278-279. These obligations naturally differ between contract types (e.g., in relational contracting where the contracting parties cooperate extendedly and in consumer contracts the obligations to inform are

*information*, that helps people to avoid harm, is different and the law typically requires informed people to disclose it as Robert Cooter and Thomas Ulen point out[788].

These rules do not establish an obligation for the vendors to provide the customers with safety information in the strict sense; i.e., the buyer has no right enforceable in a court to demand the vendor to inform her. There is no right to the vulnerability information (private information) from vendors in the sense they have a right to governmental information. The vendors must supply safety information in order to shield themselves from liability for hazardous products. Liability rules and competitive pressures in practice give the public access to much information, but without a formal right to it. The practical effect that the buyer will receive safety information related to the goods is created, e.g., in contract law by setting default rules of defectiveness of products which define the "normal" standard which the buyer can expect the goods to conform. Thus the seller has an incentive to inform the buyer of the defectiveness of the product, or face liability.[789]

This influence mechanism resembles the vertically packaged informative instrument of vulnerability reporting analysed in the previous section. Both instruments are, at the final end, directed at providing information to the users and potential purchasers of software and, by enabling them to make rational choices under fuller

---

stronger), the relationships between contracting parties (e.g. there typically are more duties to inform under consumer contract law) and regulated issues (e.g., insurance markets are typically subject to heavy duties to inform) but the tendency is clear. This is also acknowledged by Marko Mononen, *Yritysten välinen tuotevastuu*, p. 212, in relation to the duties to disclose safety information under Finnish contractual product liability rules.

[788] Cooter and Ulen, Law and Economics, p. 275-276. As noted by Janet Weiss in Public Information, p. 241-242, disclosure policies are designed to promote informed choices under conditions of some risk. The risk may be that of imperfect information as such, but usually addressed in health and safety. The disclosure rules help to mitigate risks associated with products that are legal but dangerous in some respects.

[789] Mackaay, The Public's Right to Information, p. 172; Riesenhuber, Party Autonomy and Information in the Sales Directive, p. 353.

information, at influencing the development of more secure software. However, the structure of the influence mechanism of both the legal duties to disclose and the vulnerability reporting by external parties is the same only to a degree.

Unlike in vulnerability reporting, the requirement to disclose in the case of legal duties to disclose comes from the state-actors and is based on the threat of contractual or delictual (tortious, non-contractual) liability when failing to disclose vulnerability information; vendors are liable for the non-acceptable defects in their products that the customers did not know about. Note the difference in vertical packaging. Whereas in the case of vulnerability reporting the vendor's duty to disclose in based on a threat of disclosure of information directly to the final objects of regulation (the users and the public at large), in the case legal rules the vendor's duty to disclose is based on the threat of liability[790].

> Recalling the definition of regulation above where the threat of using a regulatory instrument is just another way of employing it (implementing) and not a separate type of instrument, we can see that the former threat is an informative instrument (threat to disclose directly to the public) and the latter threat is a prescribed/written norm (threat of liability). Thus the vertical packaging in the former case is informative instruments at both levels (reporter-vendor and vendor-user) and in the latter case of prescribed norm with informative instrument.
>
> Both packages could thus be categorised as informative instruments in the classification used in this study. The instruments influencing the behaviour of the final objects of regulation (the users of the software) are informative in both of them. However, in the case of product liability rules, the classification according the lower level of the vertical package of regulatory instruments proves inapplicable[791].

---

[790] The potential of legal sanctions is an additional force towards better informing by the vendors. This holds the potential for stronger influencing capability than in the case of vulnerability reporting. However, as will be discussed below, it does not necessarily lead to stronger influence.

[791] If taken to the extreme, such classification would lead to the conclusion that liability rules are, basically, either informative instruments in encouraging

> This is why this feature of product liability rules is analysed in here as part of the written norms category, even though it so much resembles the influence mechanism of informative instruments[792].

Whereas vulnerability disclosure employed both vertical packaging and direct provision of information to the final objects of regulation, liability based duty to disclose is purely vertical packaging. No direct information provision by the regulator through its own organisation can be possible since the regulators do not posses such information. This highlights the central difference between these "informative instruments" concerning vulnerability disclosure. The regulators in the case of liability based duty to disclose do not posses the information themselves and for this reason only encourage the vendors as intermediaries to disclose and merely expect them to gather the vulnerability information[793]. In the case of vulnerability reporting the external reporters possess the information about the vulnerabilities

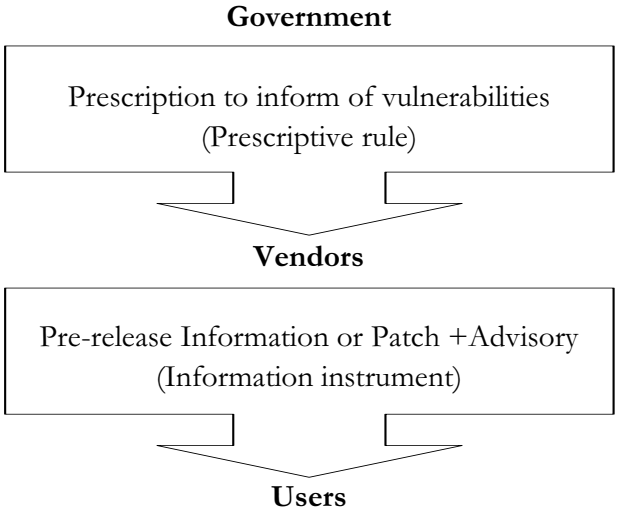
---

information provision, or technology-harnessing instruments in inciting the offering of more secure software for users. And when considering other ways in which companies can manage their product liability risks, e.g., insurance and contracting, product liability rules would have to be categorised into all of the instrument types (possibly excluding procedural).

[792] The combined analysis can be further justified by both of the influence mechanism deriving from the same rules and being subject to the same influence shaping factors. The combined analysis of the influence mechanism as written norms reduces overlap and avoids same issues from being repeated in roughly similar contexts. Another reason for combined analysis is that product liability rules do more than just provide the practical effect that software vendors have a motivation to disclose vulnerabilities in order to shield themselves from liability. Their primary objective is not to increase information provision. Instead, the intention behind these rules is to improve the incentives for enhancing the security of the goods. Thus, their primary object of regulation is the vendor and not the user. This makes the regulation different from such statutory warning and disclosure requirements that obligate vendors to provide certain standardised information to their customers (e.g., the obligation to inform about the identity of the business in e-commerce under article 5 of the Directive 2000/31/EC on electronic commerce). The product liability rules analysed here provide only the practical effect that information is provided.

[793] This is depicted in picture 5-1 below.

(act even as gatekeepers for the vulnerability information), and even though they also utilize the vendors resources not just to fix the vulnerability but also to distribute the knowledge of the vulnerability further, they also do it themselves[794].




Picture 5-1 Duty to disclose as vertical packaging

By forcing vendors to inform the users and potential buyers of the risks involved in the use of the products in question, the potential buyers are enabled to make optimal choices with closer to perfect information. The role of disclosure rules is central in the rational actor model accepted in the EU. They improve the possibilities of rational users to make right choices in the market[795]. With optimal information the buyers can operate as central controllers of the prices and quality

[794] See picture 4-1 above at page 233.

[795] This market rationality of the obligations to give information especially under EC consumer law is emphasized, e.g., by Thomas Wilhelmsson in *Social Contract Law and European Integration*, p. 128. Wilhelmsson also highlights the differences in Nordic countries. According to Wilhelmsson, market rationality is in the background of the EC consumer contract law more widely than just in the obligations to give information (idem., p. 126-148).

of products by striving rationally after their own private preferences. Potential buyers are enabled to assess the quality and security of the software in question and the claims that vendors make about them[796]. In addition to the enhancement of the market rational behaviour, and at the same time with it, disclosure rules can also be seen as protecting the informationally weaker party from entering into highly unfavourable and possibly damaging transactions[797].

At the same time, in encouraging increased vulnerability disclosure the product liability rules hold the potential for enabling the users of the software to take protective measures. For example, vulnerabilities to which users knowingly consent to while purchasing the software enable them to take protective measures against them. In the case of legal duties to disclose this applies especially to the degree duties to disclose vulnerabilities found after the release of the software are also created[798].

---

[796] Note that the stricter consumer protection rules inciting vulnerability disclosure can have significance also in relation industrial purchasing as noted by Risto Väntsi in *"Varoitanko kuluttajaa tuotteen vaaroista?"* p. 49, with reference to Jenkins JRG (1990) Consumer media as an information source for industrial products: A study, *Industrial Marketing Management*, p. 81-. Since the professional buyers are also private consumers, they also receive information directed at consumers and can be influenced by it.

[797] In other terms these instruments both reduce various forms of transaction costs (especially those stemming from the search of product characteristics) and diminish the possibilities for opportunistic (strategic) behaviour by the vendor where the misbalance of information is misused to gain undue (excessive) advantage. The transaction costs reduction is in line with the market-rationality and the diminution of possibilities for overly opportunistic behaviour resembles weaker party protection. Mackaay and Leblanc in *The Law and Economics of Good Faith in the Civil Law of Contract*, even raise the reduction of transaction costs and avoidance of opportunism as general objectives of contract law.

[798] In the Finnish doctrine the possibility of the duty to inform of the correct use the object of the contract or of risks discovered post-release is acknowledged, e.g., by Mika Hemmo in *Sopimusoikeus I* , p. 276, and in relation to consumers contracts by Väntsi in *"Varoitanko kuluttajaa tuotteen vaaroista?"*, p. 152-158 and 214-216. In Finland this influence is strengthened by the obligations to inform set for electronic communications operators in section 21 of the Act on the Protection of Privacy in Electronic Communications

## 5.3 Factors shaping the influence

### 5.3.1 Objectives

In the case of consumer product liability rules, the objectives of regulation are relatively clear. As made explicit in the drafting material of the Finnish Product Liability Act (694/1990) there are three primary objectives. The first is to improve the right of the injured party, i.e., consumers at large, to receive *compensation* from product accidents. When the consumer is provided with better means to get compensation, the economic burdens of product related accidents are distributed more fairly[799]. The second objective is to *prevent* product related accidents. The legislators considered that when the threat of liability to pay compensation is heightened, product manufacturers and importers would improve their quality assurance and monitoring mechanisms which would lead to the decrease in the number of defective products in markets. The third objective related to the EU wide harmonisation and to the improvement of the internal market (especially movement of goods) together with the provision of equal possibilities for competition for all member states.[800]

Whereas parliamentary acts typically state their objectives either as a preamble, in a specific rule about the objectives of the act, or as a direction in legislative drafting material, doctrines in contract or tort law in many cases cannot be said to have definite stated objectives[801].

---

516/2004, aka Lex Sonera, according to which also special information security risks, those that are not feasible to remove, must be disclosed to the users.

[799] The preamble of the Product Liability Directive (85/374/EEC) states even that the established liability without fault on the part of the producer is the *sole means* of adequately solving the problem of fair apportionment of the risks inherent in modern technological production.

[800] Government bill 119/1989 for Product Liability Act, p. 16. Available in Finnish at the web pages of the Finnish Parliament, http://www.eduskunta.fi/ [23.2.2006]. It is important to note that when drafting the bill, the legislator did not take information technology into consideration.

[801] As noted by Hugh Collins in *Regulating Contracts*, p. 79, this absence of explicit policy justifications is both strength and a weakness of private law

The formulation of principles in private law doctrines describes a deontic order rather than a set of instrumental goals.

The analysis of the objectives shaping the influence of a regulatory instrument on certain societal phenomenon (in this study the context of secure software development) is naturally more fit to those instruments that are explicitly intended to alter certain behaviour or to correct specific problems (be that legal on non-legal type of regulation) than those that are drafted to apply to a variety of situations. However, certain higher level objectives can be made explicit also in these situations. At least both compensatory and preventive objectives are present also in the contractual and non-contractual liability rules[802]. The objectives are typically made explicit when the private liability law rules are drafted in the form of parliamentary acts as is the case in Finland[803].

To the degree doctrines in torts conceptualizes purely in terms of corrective justice[804], i.e., tort law scholars largely consider the

---

regulation. The vagueness and ambiguity leaves potential for each generation to reinterpret the rules of private law in accordance with contemporary policy objectives. The weakness is that private law persistently evades criticism about its efficiency and efficacy as a regulatory system.

[802] Marko Mononen in *Yritysten välinen tuotevastuu*, p. 40-47 and 63, identifies both compensatory and preventive objectives in the Finnish legal doctrine under contractual and non-contractual (delictual) product liability rules. Juha Häyhä in Ankara vastuu ja vahingonkorvausoikeuden järjestelmä, p. 139, and Timo Kaisanlahti in Riskin pulverointi vahingonkorvausoikeuden tehtävänä, p. 85-105, discuss the roles of these objectives under Finnish tort law and the economic analysis of the law. Gary T. Schwartz in Mixed Theories of Tort Law discusses the role of these objectives under the common law tort doctrine.

[803] Main Finnish laws in relation to liability for damages under contract and tort law are the Tort Liability Act (412/1974) and the Sale of Goods Act (355/1987). In the drafting material of the Tort Liability Act (Government bill 187/1973, p. 8) the compensatory objective is explicitly stated. Even though the preventive objective is not made as explicit, several justifications of specific rules (e.g., liability of officials and employees) clearly imply its presence as also noted by Antti Kivivuori in Vahingonkorvausvastuun tarkoitusperät, p. 166.

[804] As Gary T. Schwartz in Mixed Theories of Tort Law, p. 1802-1811, discusses in relation to the U.S. tort law scholarship, corrective justice tends to prevail among traditional legal scholarships and the economically minded tend

primary objective of tort law being correction instead of deterrence and the theories of tort law doctrines are explained under the concepts of corrective justice, this internal argument can become a hindrance for the regulatory capacity of product liability rules, and especially of their modification to better regulate secure software development. If the concentration is purely of reparation and corrective justice arguments in the enactment, and especially interpretation, of product liability rules, there is a risk that deterrence objectives are not fulfilled and the problems of product liability rules in relation to its preventive effect are not taken seriously. Fortunately, recent argumentation especially in tort law recognises both of these objectives as relevant and considers that they can coexist and be complimentary[805].

However, the risk of over-reliance on corrective objectives still exists especially in relation to the interpretation of the vague provisions and doctrines in courts. The problem is created when court argumentation concentrates solely on facts and on issuing the burden of liability ex post[806]. If the court does not take notice of the effects

---

to favour deterrence explanations.

[805] As Schwartz, Mixed Theories of Tort Law, p. 1834, concludes his comparison between the roles of deterrence and corrective justice or retributive arguments in tort law and criminal law, "[a]s tort objectives, then, corrective justice and deterrence can be recognised as collaborators rather than competitors". As suggested by Schwartz the combination of deterrence and corrective justice can provide a better or fuller explanation of tort doctrines than can either theory standing on its own (idem., p. 1801).

[806] As Schwartz in Mixed Theories of Tort Law, p. 1815-1816, notes in developing a mixed theory of tort that acknowledges both correction and prevention as objectives of tort law, the very structure of a tort action is rooted in corrective justice. This is visible from the structural features of tort action according which liability is imposed only on those particular negligent actors whose conduct fortuitously produces injury and not on all actors that operate tortuously, and the amount of liability is the size of the particular injury rather than the value of expected risk. This complies with the logic of corrective justice but departs from the logic of deterrence.
Schwartz develops the potential for a mixed theory of tort law, i.e., the coexistence of both corrective and preventive objectives, partly on the argument that the rationality behind them is different (whereas corrective justice is concentrates on the burden of liability ex post, preventive arguments

liability rules have on prospect of liability ex ante in the decision-making of the objects of regulation, then preventive considerations and the whole objective of prevention is overlooked. This continues at least as long as corrective justice explanations are the sole means used to bring the coherence and consistency sought in legal argumentation.[807]

A severe hindrance to the regulatory capacity of product liability rules, and especially their development towards better regulation of secure software development, arises also from the conflicting interest surrounding the issue. Whereas the objectives of the representatives of consumers are towards enhancing the possibilities to get compensation in practice and not just the right to compensation[808], the objectives of the producers are towards maintaining status quo,

---

concentrate on the prospect of liability ex ante). I have no quarrel with that. I am not arguing against the coexistence of both objectives. Instead, my argument concentrates on the use of both corrective and preventive arguments in courts. If courts do not anticipate the preventive measures taken by the defendant group, correct incentives are not given.

[807] Even though this idealized argument does not fully depict the reality of court argumentation in tort cases, there is certain truth behind it. As Jane Stapleton, Regulating Torts, p. 125, notes in discussing Hugh Collins' arguments favouring regulatory analysis of the effectiveness of private law in common law, many of the legal concerns confronted with courts in tort cases cannot be described in terms of consequentialist policy (a regulatory goal). Consequential concerns do play a role in court argumentation, but the emphasis is on doctrinal arguments.

[808] An example of such situation can be found from the Finnish implementation of the Product Liability Directive. Even though deterrence objective was made explicit in the implementation of the Product Liability Directive in Finland, compensation was considered to be the primary objective. However, the Finnish Product Liability Act concentrated only on enhancing the *right* of the injured party to get compensation. As Thomas Wilhelmsson and Matti Rudanko in *Tuotevastuu*, p. 35, point out, no such arrangements that would have guaranteed the compensation *in practice* were established. For example, Finnish legislator followed the lines of the Directive and did not make product liability insurance obligatory. The inexistence of mandatory product liability insurance has continued to be raised by consumer representatives as a reason why the Product Liability Directive (85/374/EEC) does not adequately protect consumers (Meltzer et al., *Product Liability in the European Union*, p. 42).

limiting the right to liability, and minimizing possibilities to actually receive compensation[809]. This conflict of interest hindered the approval of the Product Liability Directive (85/374/EEC) and its implementation in Member States, and still continues to hinder the development of product liability rules towards more effective regulation.

The objectives of correction and deterrence are common to objectives of liability rules in general. They concern the variety of cases to which the liability provisions apply. Of the laws and doctrines themselves, no specific objectives in relation to specific policy areas such as secure software development can be identified. The goal that product liability regulation under special, contract or tort law pursues naturally is not directly secure software development. However, due to wide application of the rules to multiple occasions, it can be used to tackle also the societal problem of insecure software. The above mentioned strength of private law regulation in relation to the absence of explicit policy justification enables the rules to be interpreted in accordance with the objective of secure software development[810].

---

[809] This is visible in the opinions of both consumers and producers in relation to the balance stroke by the Product Liability Directive. Even though 66% of all those participants who expressed their view in an EU-study (Meltzer et al., *Product Liability in the European Union*, p. 42) on the need to reform the Directive thought that the Directive did strike an appropriate balance, only 20% of the consumer representatives thought so. 80% of them though that the Directive did not adequately protect the needs of the consumer, whereas 26% of the producers though that the Directive does not adequately protect them. A similar pattern of responses emerged in relation to the product liability system as a whole, defined as the combination of the Directive, national laws and procedural rules (Meltzer et al., *Product Liability in the European Union*, p. 43). The conflicting interest are present especially in the discussion about the need to reform the Product Liability Directive as the study of John Meltzer, Rod Freeman, and Siobhan Thompson, *Product Liability in the European Union*, p. 46-52, testify especially in relation to the burden of proof and the development risk defence.

[810] Objectives are formed when the doctrines are applied to specific factual situations in court cases. The policy objective of liability rules can then be found from the purposes to which the liability doctrines are used. Thus, they could be used to influence secure software development. Since the court cases are lacking, no specific context based policy objective have formulated in

One of the goals of these general private law liability rules can be formulated to achieve a situation where citizens, consumers, and organisation can use non-damaging products and know the possible risks involved in the use of the products that contain acceptable vulnerabilities that are not feasible to remove (e.g., in the case where the user is in a more cost-efficient situation to take precaution). In this way it also contributes to the general quality and safety of products. Thus, also the security of software could be formulated into the background of product liability rules even though their objective is more general (i.e., related to the overall quality of all products and safety of users).

The objectives of those drafting the software contract templates become essential due to the extensive dispositive nature of the legal duties to disclose, i.e., the rules establishing the duties to disclose can be contracted otherwise. Fixed standard software templates are widely used in COTS business, instead of widely negotiable templates characterising tailored and MOTS software businesses, especially in the end-user licensing[811]. These fixed templates used reflect a major legal source according to which the standards for product liability are evaluated. The contractual practices visible in them override the dispositive default private law rules as legal sources.

When the templates are drafted by one of the contracting parties, they typically represent the objectives of the enterprise that drafted them[812]. This raises the opportunity to strive for better legal status compared to the weaker contracting party who basically faces the option of accepting the terms or not making the contract. The

---

relation to secure software development and the discussion in here is pure speculation of the writer.

[811] Warsta, *Contracting in Software Business*, p. 182.

[812] This might be the case in many of the end-user license agreements directed at consumers, even though no generalization can be made. Whether or not behaviour is opportunistic and against good faith doctrines depends on case by case decision-making. However, when either side of the contracting parties is drafting the terms alone, it should be kept in mind that the expectation of opportunistic behaviour is more easily made.

situation might be the same even when the standard terms are drafted by an industrial organisation representing only either of the contracting parties, with the nuance of meaning that the branch, instead the single firm, might strive for a better legal status. Even when agreed documents are used that are drafted bilaterally by the representatives of both contracting parties or even by all representatives of the industry multilaterally, it does not necessarily mean that the interests of both parties are considered. The power relations and expertise of parties define the content of the agreed documents to a degree.[813]

Disparity of wealth especially in the case of consumers and the possible strong market position of the COTS software vendor also partly explain the use of such clauses. This can influence the license terms in ways that make them less favourable for the users; especially when the customers do not have the capacity (information and expertise) to negotiate above the standard contracts. However, this unequal bargaining power argument does not fully explain the use of standard form contracts with extensive liability disclaimers and exclusion clauses in the COTS software business. Standard form contracts are used more widely in software business than the theory of inequality of bargaining power suggests. Similar standard form contracts are also used between businesses with relatively equal bargaining power.

> The normal legal presumptions of approximate equality of bargaining power and comparable sophistication in evaluating benefits and risks might not be correct in the context of software security. A typical purchaser (even industrial), excluding only the largest industrial buyers that have the resources to use the expensive and scarce expertise, is only slightly more experienced than a consumer and should not thus be treated as having sufficient information for rational decision-making on costs of software vulnerabilities[814]. The unequal bargaining

[813] Wilhelmsson, *Vakiosopimus*, p. 31-32.

[814] Even though industrial buyers are in a better informational position to evaluate the products and vendors claims than consumers, they still are at the mercy of the vendor largely due to the lack the means (not only are white-box methods largely unavailable, but also the tools for testing even on the

> power argument explains the use of extensive liability disclaimers and exclusion clauses in standard form contracts only partially. It might be correct in relation to software security and other credence characteristics of software, but not in relation to other quality aspects.

In addition to the superior bargaining power, also the potential to exploit the market failure for contract terms raises the opportunity to strive for better legal status from that provided by the default terms. Consumers are unlikely to extend their search for goods beyond price and quality comparisons. In most cases buyers are unlikely to analyse standard form contracts in order to make comparisons. Instead, the software contract (titled typically as a license) is approved together with the installation of the software product without further consideration of the meaning of the contract terms (outside possible anecdotal knowledge of their total exclusion of liability). Vendors can then take advantage of this market failure by promulgating a standard form containing small print exclusions and disclaimers of liability, which, if interpreted literally, almost relieve them of any obligations under the contract. Even though this can happen even between businesses when there is no time to analyse the meaning of contractual clauses in detail, not even the software business as a whole is likely to be that hasty.

As Hugh Collins notes, standard form contracts certainly reduce transaction costs and are useful in permitting organisations to regulate their business relations in ways that take into account the peculiarities of the business[815]. The transaction cost reducing explanation combined with the efficient governance structure argument is the most pervasive reason why standard form contracts are used. However, it has to be kept in mind that it is only one of the reasons.

Following Collins, mass contracts provide an efficient governance structure[816]. They permit organisations to regulate their business

---

developer side are relatively undeveloped and not in wide use) and financial resources to evaluate the products security (high costs involved).

[815] Collins, *Regulating Contracts*, p. 232.

[816] Collins, *Regulating Contracts*, p. 232.

relations in ways which economize on transaction costs. Even though standard form contracts also might be an expression of exploitation of market failures for contract terms or unequal bargaining power, the starting point is that there are valid reasons for their use. They become objectionable only when in the context of the absence of a market for contract terms and the absence of competition for fair complaint mechanism[817]. In these instances, it becomes objectionable, because the standard form is used to redistribute power relations in ways which permit oppressive post-breach bargaining situations.

### 5.3.2 Substance

Due to the variety of norms on which product liability rules derive from, the substance of product liability regulation can be manifold. The various laws and court cases in contract, tort or specific product liability rules that establish quality and security standards in law are primarily voluntary and supplement standards in contracts. These model set of rules provide the detail of the regulation in the absence of express contractual terms. Mandatory standards define the substance of product liability in consumer protection legislation.

> Due to the huge variety of rules that set the substance of product liability rules[818] and their large contextual variation under different types of contracts, the influence of the different standards are not analysed in detail. However, the centrality of the different standards for the influence that the product liability rules can provide for the behaviour of the software developers and the users of the software in question must be acknowledged. Different bases for liability under the various rules (negligence or strict liability with several differing excuses and defences), which differ even in the partially harmonised EU products liability regime, can produce a huge variety of possible

[817] Collins, *Regulating Contracts*, p. 232.

[818] Actually most of the contract and tort law rules could be applied to liability for software defects as the year 2000 problem showed (Hemmo, Vuoden 2000 ongelma ja siviilioikeus [The year 2000 problem and civil law], p. 18).

effects as economic analysis of the law has shown[819]. However, when even the mass of theoretical economic research cannot fully agree on the effects of various liability rules and empirical evidence on them is scarce, as Michelle J. White points out while trying to empirically test both comparative and contributory negligence rules in U.S. accident law, a detailed analysis of several liability rules simultaneously would become too laborious for a study where the purpose is not to analyse the effects of liability rules as such[820].

For the same reasons the varying remedies available for plaintiffs in tort and contract law actions, such as monetary damages, and specific performances like repair of defective good, reduction of price, withdrawal of a contract, are not analysed separately even though their use influences behaviour in different ways as economic theory has shown[821]. Of the different remedies available in product liability cases, the emphasis in here is primarily on the liability for damages also in the case of a breach of a contract. The differences between alternative legal traditions might be greater in theory than in practice due to different legal systems of law responding to the same economic logic as Robert Cooter and Thomas Ulen suggest[822], but variations

---

[819] A short overview on the incentives for precaution and for the amount of doing the potentially harmful activity created under various bases of liability under tort law is provided by Robert Cooter and Thomas Ulen in *Law and Economics*, p. 300-313. A short overview of the wide U.S. literature is provided by Daughety and Reinganum in Product safety, p. 1191-1193.

[820] White, An Empirical Test of the Comparative and Contributory Negligence Rules in Accident Law, p. 308-309. Such an analysis would be too complicated and heavy since a detailed analysis would require going through all the various rules and their associated excuses and defences together with their damage calculations rules. Such an analysis would be feasible only if we were searching for the most efficient rules. Since this is not the purpose of this research, the analysis concentrates on the capabilities of liability rules to influence secure software development, the detailed analysis of the ways various liability rules influence behaviour can be left for further studies. Only the possibility of liability rules to influence secure software development is analysed. A basic understanding of the ways liability rules influence behaviour is sufficient without a deep analysis of all the possible variants. Analysis concentrates on liability law as a system.

[821] As Cooter and Ulen, *Law and Economics*, p. 226, note in relation to contract law, different remedies create different incentives for the parties to a contract.

[822] Cooter and Ulen, *Law and Economics*, p. 226.

> are still so large that simultaneous analysis in a study concentrating on the capabilities to influence rather than on effects and efficiency alone is overly burdensome and would require reliance on original application of economic research methods to a non-feasible degree.

Contracts make explicit in most detailed way the substance of the product liability rules. They are the primary legal source in relation to COTS software. However, a regulatory analysis has to start from the dispositive default rules. In order to understand the efficacy of regulation by liability rules it is not possible to bypass the initial legal positions and default rules of liability made explicit by legislators and long line of court praxis. It is these default rules that the contracts typically strive to alter. Understanding the role of the standards set in contracts is not possible without first analysing the capabilities of default allocation of risks in typically dispositive default contractual or tortious liability rules in regulating behaviour.

*Applicability to software.* The default rules that form the background of product liability contain several uncertainties[823]. Already the coverage of product liability rules is problematic in relation to software. There is no certainty of whether especially the liability rules for defective products in general contract and specific product liability laws as such apply to any type of software. Without going deeper into the several and widely discussed uncertainties, that still remain largely unsolved in many jurisdictions, like whether and which kind of software is a good/product as defined under these norms, and whether the contract for software purchases is a license for use or sale where the ownership is transferred (critical for the applicability of the sale of good provision), a basic starting-point that they do apply to COTS

[823] The openness of the interpretations of the applicability of sale of goods and specific consumer product liability rules, the confusion about the relationship between license contracts and contracts of sale and the differences in the treatment of COTS, MOTS and tailored software, are still present in many jurisdictions even though some progress has been made even in Finland since 1999 when Mika Hemmo in Vuoden 2000 ongelma ja siviilioikeus [The year 2000 problem and civil law], p. 17, emphasized the confusion in relation to the Y2K problem.

software is accepted[824].

Of the types of market driven embedded and COTS software that are the subject of this study, the applicability of product liability rules is unquestionable in relation to embedded software[825]. In relation to COTS software in general and especially other software than operating systems (application software) there still are open questions especially in relation to the license for use and contract for sale distinction that is considered necessary for the application of contractual and specific product liability rules. However, the tendency seems to be towards wider application of both the sale of goods and product liability rules. COTS software is increasingly being considered as a good. Traditional reluctance of accepting the provision of software as a sale of goods by stating that it actually is a supply of service or license for use, which are not covered by product liability rules in general contract and specific product liability rules, is vanishing in relation to COTS software.[826]

Note that in Finland, recent commentaries on product liability still seem to favour the interpretation that the rules in the Finnish Product Liability Act (694/1990) apply to software that is delivered in a cd- or a dvd-rom, or is bought together with the computer containing the software, but do not apply to software that is delivered over networks or is installed by the supplier during a visit in the customer's premises[827]. Such an interpretation leads to a spurious division in the liability for software defects. The mode of delivery of software ought not to be the decisive factor in considering the standard by which

---

[824] For wider discussions about the complexities of applying liability rules in the information technology sector in general see, e.g., the analyses of Mats Bryde Andersen in *IT-retten*, especially chapter 18, and in *Edb og ansvar*.

[825] The applicability has been typically accepted also in cases where the software is sold as part of hardware (e.g., in a cd-rom), is necessary for the functioning of the hardware (e.g., operating systems), or used to control some other product (embedded software).

[826] For a recent and a wider overview of these issues see, Lloyd, *Information Technology Law*, chapters 26-27 (concentrating on the law of United Kingdom).

[827] Mononen, *Yritysten välinen tuotevastuu*, p. 104–105 and 107; Wilhelmsson and Rudanko, *Tuotevastuu*, p. 79–81.

the software will be judged (e.g., whether it is some type of a satisfactory quality/fitness for purpose test in contractual liability law for tangible items, or some type of a statutory requirement on the supplier to use reasonable skill and care as in the contractual liability for the supply of services). Neither should the mode of delivery impact the possibility of excluding liability for failing to meet that standard[828]. At least it ought not to weigh more than the business model in use or the way development is done. However, this peculiar situation exists in the private law of a number of European countries (e.g., Belgium, Denmark, Italy, Spain and the UK) beside Finland and even more generally than just in product liability law[829].

It is not necessary to go deeper into this extensive and still largely unresolved discussion about the final applicability of sale of goods and product liability rules on software. For the purposes of this study the important thing is the influence of this uncertainty in the application of the rules to software to the incentives these rules provide. This is discussed later on.

When the basic starting point that product liability rules both in consumer and enterprise relationships apply to the COTS software vendors in general similar to producers of every types of goods is accepted, it becomes feasible to study the liability as regulatory instruments in relation to secure software. You do not even need to accept this basic starting point, i.e., that definitions of a defect in sale of goods and product liability laws directly apply to software, in order see the usefulness of such a regulatory study. There is a strong argument that these rules provide analogical value when considering the qualitative requirements of software. For example, the rules on sale of goods are considered as basic general principles of all types

---

[828] In consumer contracts for tangible items it typically is not possible, whereas for contracts concerning the supply of service the exclusion of liability is more easily allowed.

[829] This has been pointed out in 2005 by two practitioners of UK law, Dominic Callaghan and Carol O'Sullivan in Who Should Bear the Cost of Software Bugs?, p. 56.

of contracts and thus provide guidance[830].

*Damages covered.* Even when the application of product liability rules to COTS software is accepted, the scope of the rules still limit the regulatory capability of these rules towards secure software development. As already pointed out, the consumer product liability rules based on the EC Directive (85/374/EEC) are limited only to death, personal injuries, or damages to property (other than the defective product itself) that is intended and mainly used for private use or consumption[831]. Thus damages to other (non-consumer use) property and pure economic losses that have no relation to damages to person or property are not covered[832]. When the largest damages that result from vulnerabilities are typically economic in nature (e.g., loss of income due to downtime in service caused by an attack abusing vulnerability in the software used) and are directed towards economic property such as computer equipment, most cases of damages resulting from misuse of security-related vulnerabilities are excluded under the directive[833]. However, under tort and contract law also

---

[830] Of such argumentation in Finland, see Takki, Vuosi 2000 –tietojärjestelmät ja vastuut yrityksen näkökulmasta, p. 37, footnote 10; and in Sweden, see Lindberg and Westman, *Praktisk IT-rätt*, p. 271. Similarly also in Bradgate, Beyond the Millennium, heading 4, first paragraph, in relation to the laws of the United Kingdom.

[831] See page 298 above.

[832] As Mathias Reimann notes in Product Liability in a Global Context, p. 150-151, pure economic loss is not available in strict product liability in most countries.

[833] When only damages to consumer property are covered, the importance of the Product Liability Directive and implementing legislation is typically limited only to damages to persons. For example, if embedded software (in a medical device or a complex tool) causes physical damage to an employee, which in turn results in costs for the employer, can the Directive be applicable also in commercial relations. However, these cases are typically covered by accident insurances and compensation is sought from them instead of product liability. In these cases the regulatory capacity of specific product liability laws are of course stronger. But this does not change the situation for the main stream COTS software vendor whose products are not used in such critical functions that can lead to damages to persons.

damages to commercial property are typically covered, excluding the sale of consumer goods laws[834], and in contract law also the economic losses are compensated relatively widely with extensive national differences even in Europe.

It is important to note that not all damages resulting from defective (insecure) software are purely financial in nature and thus have limited recoverability especially under the Product Liability Directive (85/374/EEC). To the degree COTS software and data as such are accepted to be products also under the Directive, the damages more easily have a relation to property; also to other property than the defective software itself[835]. This is the case, for example, when defective software allows a virus or a worm to erase the user's hard drive.

In these cases the division between direct and consequential (indirect) damages becomes relevant in contractual relations[836], since the economic damages resulting from the defective software typically have a relation to property (i.e., the software and the data in the hard drive). Direct damages like the value of the lost data, costs of restoring from a backup copy or otherwise replacing it, the decrease in the value of the data (e.g. due to its lost reliability) or the time directly related

---

[834] The rules and presumptions of the conformity of consumer goods with the contracts has been minimum-harmonised by Directive 1999/44/EC of the European Parliament and of the Council of 25 May 1999 on certain aspects of the sale of consumer goods and associated guarantees, OJ L 171, 7.7.1999, pp. 12-16.

[835] Note that damage to files and data inside a computer does not fall into the property damages category without difficulties. However, damage to data inside a computer has been compensated as property damage for example in Denmark, as reported by Egeskov and Christensen in Behovet for forsikring af softwareproducentens mulige erstatningansvar, p. 59, in 2003. This is noted also by Thomas Wilhelmsson and Matti Rudanko in *Tuotevastuu*, p. 231.

[836] At least in the Finnish doctrine, the division between direct and consequential damages has less significance in tortious (non-contractual) relations. As pointed out by Wilhelmsson and Rudanko in *Tuotevastuu*, p. 214, consequential damages are compensated with similar conditions as direct damages as long as they have a relation to damage to persons or property.

to the error correction are more easily recoverable[837]. However, they are typically relatively insignificant when compared to the loss of income due to the downtime of the service or other consequential damages like poor performance due to IT trouble and loss of confidence towards the IT system, loss of image or the reduction of the stock price of the company due to the loss of personal data being published[838].

*Vague standard.* In the formulation of standards for guiding participants in markets, private law regulation in general suffers from the structural weakness of setting standards with an adequate degree of specificity in order to provide effective guidance. As is typical for private law regulation, the default rules and compulsory terms about product

---

[837] In Finland, for example, the liability for *direct damages* in contractual situations is typically based on control liability, which is a non-negligence based liability where the vendor has the burden of proof in relation exemptions of liability; i.e., the vendor has to show that the cause for the defect was outside its sphere of influence and that it could not have been required to consider the defect when drafting the contract within reasonable limits, or that it could not have avoided or overcame the consequences, in order to be exempted from liability. Liability for *indirect (consequential) damages* is, instead, typically negligent based. There are, however, in the Finnish contractual liability doctrine rules on the burden of proof that bring the liability for indirect damages close to the control liability applied to direct damages. For example, presumptive liability, which is culpability (liability based on negligence) based on reversed burden of proof is common in contractual relations; in order to be exempted from liability the vendor has to show that it has acted diligently or, alternatively, that it has not been negligent. The vendor can under presumptive liability rule also plead to certain causes of defects that are in its sphere of influence, but the results of court decision rarely differ.

[838] Even though such losses can be recoverable under general law of torts, it is important to note that under the Product Liability Directive (85/374/EC) they are not. When the Directive allows the recovery only for damages to consumer use property (damages to other property than the defective product itself that is intended and mainly used for private use or consumption) the losses from downtime of business activity caused by damage to business assets do not fall under the scope of the directive. Under the Directive, relevant pure financial losses could appear, e.g., when an existing vulnerability that not yet has caused any damage has to be repaired. Similarly in Wilhelmsson and Rudanko, *Tuotevastuu*, p. 212.

liability describe open-textured and general standards. These cannot inform participants in market of what is required, so they prove an inapt tool of guidance[839]. This is so especially in relation to the required security level of products.

Even though the lack of specificity diminishes the directive force of private law rules, it must be noted that their open-textured formulation is a necessity. This is due to the application of private law rules to various and changing circumstances. The use of abstract and general rules enables private law to persist over long periods of time without major changes. Via interpretation the rules can be applied to changing policy objectives.[840]

The standard is made most explicit in specific product liability rules such as the EC Directive on Product Liability (85/374/EEC)[841]. According to article 6 of the Directive a product is considered to be defective "when it does not provide the safety which a person is entitled to expect". The basis for liability is thus insufficient safety and security which are defined according to objective expectations test where all circumstances are taken into account[842]. More specificity

---

[839] This is emphasised by Hugh Collins in *Regulating Contracts*, p. 293-294, in relation the contractual regulation of quality. Of the benefits and disadvantages of openly formulated rules in general, see Tala, *Lakien vaikutukset* [The Effects of Legislation], p. 188-196.

[840] As Meltzer et al., *Product Liability in the European Union*, p. 49, note in discussing the possibility of more precise definition of the concept of "defect" in EC Product Liability Directive "it is better not to attempt to define the concept with too much precision, not least because this could restrict the ability of judges to deal with matters on a case-by-case basis."

[841] Council Directive 85/374/EEC of 25 July 1985 on the approximation of laws, regulations and administrative provision in the Member States concerning liability for defective products, OJ L 210, 7.8.1985 pp. 29-33, amended by directive 1999/34/EC OJ L 141, 4.6.1999, pp. 20-21.

[842] This is not the only way to define defectiveness. The external test used to decide whether or not a product is safe enough that prevails in the United States is called risk-utility analysis. It renders a product defective if its risks outweigh its utilities. The question then is whether the product could have been made safer without unduly impairing its utility. The feasibility of a safer alternative design thus becomes crucial. As Mathias Reimann in Product Liability in a Global Context, p. 139-140, notes, the vast majority of

is provided by the examples of the circumstances that, among other things, have to be considered. These include also the presentation of the product, which brings the duties to disclose explicitly into the standard as noted above, its expected usage, and the time when it was put into circulation. However, the standard still remains relatively vague and open to interpretations[843].

The criterion for defectiveness in the Directive ("lack of safety") differs substantially from the "lack of fitness for purpose or use" and other similar formulation that play a role in determining a breach of contractual obligations in the case of defective products[844]. Whereas the criterion of fitness for purpose and other similar constructions

---

jurisdictions with a special product liability regime are in line with the European model and thus rely on reasonable consumer expectations rather than employ a cost-benefit analysis. According to Reimann, Product Liability in a Global Context, p. 140-141, most jurisdictions in the world have followed the approach that *once* dominated in the United States (consumer expectations test under the Second Restatement of Torts in 1966 which was later changed into a more differentiated cost-benefit like approach due to the influence of the Law and Economics schools in the Third Restatement) and is *now* an crucial element of the European model.
This refinement to the definition of defectiveness in the U.S. doctrine has not, however, gone unnoticed in Europe. Questions concerning whether or not it is appropriate for the courts to undertake a risk/benefit analysis when assessing what a person is entitled to expect, and the extent to which the actual conduct of a producer (such as the degree of care taken, or not taken) is relevant in this context, have arisen in case law but have not yet been finally resolved by the courts in any EU Member State (Meltzer et al., *Product Liability in the European Union*, p. 48). As Wilhelmsson and Rudanko, *Tuotevastuu*, p. 158, point with references to U.K. and German legal literature, these tests might differ only in style rather than in content.

[843]This became evident also in relation to the Millennium Bug as is visible in the many articles and presentation. See, e.g., Geraint Howells, The Millennium Bug and Product Liability, p. 300-3001. Meltzer et al., *Product Liability in the European Union*, p. 48, also refer to the incapability of precisely defining "defectiveness" under the EC Product Liability Directive (85/374/EEC) and note the similar problems in relation to the concepts of "negligence" and "fault" that have always escaped precise definition under national liability systems.

[844] This clarification is made explicit in the preamble of the Product Liability Directive (85/374/EEC).

rely heavily on the terms of the contract between parties, in product liability law the lack of safety is considered on objective grounds. This means that even if a product lacks in safety or security, it necessarily is not defective under contractual rules. Defectiveness in contract law neither necessarily means that a product is insecure or not safe. Liability is based on whether or not it is fit for purpose and insecurity of a product does not necessarily mean that fitness for purpose is lacking. Security and safety is not given a similar role in contractual liability as it is in delictual product liability.

This emphasises the importance of the safety expectations of the consumers in relation to the possible personal injuries or damages to private property resulting from defective products. It does not, however, mean that the criterion of fitness for use in the default rules on sale of goods could not establish expectations for the safety of the products. There still are in member state laws doctrines in contract that create possibilities for injured parties to claim damages in similar situation than those covered by the EC Product Liability Directive (1985/374/EC).

Even though the European Product Liability Directive (85/374/EEC) primarily relates to tort law, it affects contract law to the extent that national law has previously dealt with product liability problems as part of contract law, and the rules of the directive are mandatory under contract law[845]. However, the directive harmonises the national rules only to a certain degree. The option of using other liability systems beside the Directive is explicitly stated in article 13. The Directive does not seek full harmonisation and national systems are allowed to the degree they do not depart from the terms of the directive by imposing higher obligations for vendors or by creating higher levels of protection for consumers[846].

---

[845] According to the article 12 the liability arising from the Directive may not, in relation to the injured party, be limited or excluded by contract. Of this see Wilhelmsson, *Social Contract Law and European Integration*, p. 57.

[846] This "maximal harmonisation" approach has been confirmed in recent ECJ cases Mariá Victoria Gonzáles Sánchez v Medicina Asturiana SA, Case C-183/00 [2002]; Commission v France, Case C-52/00 [2002]; and Commission

Under these considerations it is not surprising that, despite the harmonisation by the Product Liability directive, there still are in each member state systems of extra-contractual liability for the recovery of compensation by injured persons. In almost all member states the law of contract plays a role, with significant national variations, in the contractual remedies available and types of damages recoverable in product liability cases.[847]

General contractual liability doctrines for defective products especially in consumer contracts, but also in commercial contracts, at least implicitly also assume the criterion of safety. Even when overlooking the relatively obvious cases where the contracts explicitly or by interpretation provide that the software must be free from security related defects to a certain degree, or where information on the security (especially quality aspect) of the software is given in the marketing or other pre-purchase information[848], when the breach of contract can be analysed in relation to these explicit stipulations and information provided, the default rules on defectiveness in contract law apply to vulnerable products to the degree users in general can expect certain quality and security from them. However, the basis for liability, circumstances in which contractual remedies are available even for the consumer, and the types of damages recoverable, vary significantly between Member States[849].

> As Meltzer, Freeman, and Thomson summarise in gathering the reports of national experts in product liability, in most Member States there is generally a requirement of fault or an element of "bad faith" that gives rise to contractual liability, but in countries like UK and Ireland the basis for contractual liability is strict in the sense that there is no requirement that the breach be attributable to the fault of the

---

v Greece, Case C-154/00 [2002].

[847] This has been pointed out by Meltzer et al., *Product Liability in the European Union*, p. 15-16.

[848] These situations will be discussed later on in relation to the contractual practice in the COTS software business.

[849] Meltzer et al., *Product Liability in the European Union*, p. 16-18.

> defendant[850].
> In most civil law jurisdictions in the EU, the seller will be liable only where she knew, or should have known, that the product was not in conformity with the contract. Even the extent of knowledge expected from a seller does vary between Member States. For example, courts in France and Luxembourg usually expect a professional seller to be aware of all defects, while a seller in Austria or Germany is not necessarily under a general obligation to examine, and thus be aware of any defects in the products she sells. In Nordic countries the liability for unsafe products has evolved firstly through legal usage and mainly through non-contractual (tort) law rules and later through special product liability acts. However, the contractual liability rules still are applicable to product liability cases, even though their role is limited in the Nordic countries and subject to significant national differences even in the relatively homogenous Nordic private law doctrines[851].
> Even though in most EU countries consumers can generally recover damages for personal injury or harm to property other than to the product itself, in Denmark, for example, compensation for damage to property other than to the product itself, or to a product in which it is incorporated, is usually not recoverable under contract law. In Finland and Sweden, such compensation is recoverable only in limited circumstances[852].

In non-contractual (tortious) liability rules where the standard for conduct is often explicated in the form of a negligence rule the vagueness is even larger[853]. For example, when the required level of

---

[850] Meltzer, Freeman, and Thomson, *Product Liability in the European Union*, p. 17-18.

[851] This has been shown by Thomas Wilhelmsson in Köprätten och produktansvaret, p. 627-642, and is noted also in a recent EU-study on product liability by Meltzer et al., *Product Liability in the European Union*, p. 16.

[852] This has been noted also by Meltzer et al. in *Product Liability in the European Union*, p. 18.

[853] Almost all EU Member States provide a mechanism under a system of non-contractual (tort) liability by which injured persons, who can prove that they were injured as a result of the neglect of the producer, may be able to recover most or all of their material damages and, in most cases, recover non-material damages as well (Meltzer et al., *Product Liability in the European Union*, p. 19). The

care is explicated in such vague terms as in the Finnish Tort Liability Act (412/1974) chapter 2 section 1 subsection 1 ("A person who deliberately or negligently causes injury or damage to another shall be liable for damages, unless otherwise follows from the provisions of this Act") the factors that need to be considered in order to define the level of care are even less explicit and wider than in general contract or consumer contract law or in consumer product liability laws. Whereas the traditional fault-based principles of liability in most jurisdictions focus on the conduct of the defendant, the principle of liability under the Product Liability Directive and typically also under contract law focuses on the characteristics of the product, and specifically whether the product contained a defect that caused the damage[854].

As Collins points out in relation to contract law, the standards provided by private law regulation can become more specific when applied by private ordering (contracting) and adjudication[855]. For example, guidance to the interpretation of the concept of "defect" in the EC Product Liability Directive (85/374/EEC) could be provided by increasing experience of use of the Directive in litigation and consequent emerging body of case law as expected by John

---

extra-contractual liability rules define the liability for damages in cases where there is no contract between the injurer and the injured party. As the EU study conducted by John Meltzer, Rod Freeman and Siobhan Thomson, *Product Liability in the European Union*, p. 19, show, in most Member States tort law requires that the defendant is at fault, or in breach of some general duty to the claimant. In some jurisdictions, this element is described in terms of unlawfulness or culpability. In others, it is understood in terms of a breach of a duty of care.

[854] The importance of this distinction has been raised by Meltzer et al. in *Product Liability in the European Union*, p. 48. Even though this difference might seem large at first, when the EC Product Liability Directive (85/374/EEC) refers to the expectations of a person and allows development risk defence, negligence like considerations creep in and the directive only purports to impose true strict liability in principle as Mathias Reimann notes in Product Liability in a Global Context, p. 141-142, on the basis of an analysis of product liability rules worldwide published in 2003.

[855] Collins, *Regulating Contracts*, p. 59.

Meltzer, Rod Freeman and Siobhan Thomson in their report for the European Commission on the updating of the Directive in 2003[856].

Even if such court practice would evolve, it could not serve as a sufficient guide for the market actors. As Collins notes, the standards produced by case-law tend to be tied so closely to the facts of a particular case that it is difficult to generalise their application to other cases[857]. The standard now becomes so detailed in its application, as is also the case with specific product related contracts that they cannot serve as a guide to behaviour in general. Even though it could be possible to obtain guidance by analogical reasoning from decided cases, this typically would require the use of an expensive expert lawyer.[858]

---

[856] Meltzer et al., *Product Liability in the European Union*, p. 49. They also expect that some aspects of the concept of "defect" will come to be clarified in due course by the European Court of Justice.

[857] Collins, *Regulating Contracts*, p. 81. This is especially due to the self-perception of private law adjudication being that it concerns the resolution of a dispute rather than the regulation of a market by setting a standard (Collins, *Regulating Contracts*, p 81). Other cases with slightly different facts cannot be directly considered to receive the same treatment in a court of law.

[858] There is also the danger that the general formulation of a rule can lead to dissimilar interpretations in courts and in the practice of the objects of regulation. This causes further problems in analysing the standard of care required under the law. A further hindrance for the guiding function of the private law standards has been stated to result from the incapacity of private law to publicise its standards. According to Hugh Collins, *Regulating Contracts*, p. 81, "[t]he detailed regulatory standards produces by private law can only be gleaned from the reports of court's decisions, which are seldom complete. These decisions are then only disseminated by uncoordinated private publicity in the form of books and articles about the law". However, this analysis neglects one of the basic arguments made in legal informatics; the use of information networks has already changed the publication of standards also in private law. The relative number of cases which come to the attention of the legal professional through legal dogmatics (i.e., description and prescription of the law) is diminishing due to the use data systems as Ahti Saarenpää, Court Decisions as the Focus of Study, p. 128, pointed out at the dawn of legal informatics in Finland in 1984. For Saarenpää this is a regrettable development because the only information available in the networks is a bulletin formulated by the court itself or some other similar brief commentary. As Saarenpää puts it Court Decisions as the Focus of Study, p. 150: "Researchers who use such

Further specificity for these general standards in contract, tort and specific product liability rules is provided in the existing preparatory materials[859], follow-up studies and academic research. As Wilhelmsson and Rudanko point out in relation to the Product Liability Directive (1985/374/EC) and its Finnish implementation (Product Liability Act 694/1990), this material gives relatively detailed guidelines for the consideration of the security level required by the law[860]. However, the guidelines for the consideration of the variety of factors relevant to the specification of the security standard are so diversified and complex, and the material to be analysed so extensive, that utilising them in the security considerations of a specific product still requires the use of scarce and expensive legal experts that understand the specificities of the software industry[861].

This means that even though the product liability rules inform both the vendors and the users that the good must be as secure as the users in general can expect (reasonable secure in their ordinary uses) or alike, any further detail requires the use of expensive advice of expert lawyers from the specific national jurisdictions in question to gather any useful guide about the requirements of the standard[862]. And still, the meaning of this requirement in practice remains for the courts in isolated decisions to specify in the final end. The objects of regulation have difficulties in recognising and anticipating the

---

cases easily depart from *following rules* – instead, they only follow cases. The essential element in legal activity is bypassed". [Emphasis in the original]

[859] For general contract and tort law rules the preparatory materials can be relatively uninformative for specific cases, even if they exist, due to their generality and age.

[860] Wilhelmsson and Rudanko, *Tuotevastuu*, p. 157.

[861] Relevant factors can include, in addition to the expected usage of the products and the information provided in combination with the product and its use, issues like the expectations of the users, the level of security in the relevant markets, applicable governmental orders and relevant standards.

[862] Hugh Collins, *Regulating Contracts*, p. 81, emphasises this in the UK national context and in relation to the contract law.

requirements set for their behaviour[863].

The specificity of the guidance provided by liability rules is further hindered by the various legal categories with differing backgrounds that form the basis for liability. When the liability can be based on contract, tort and specific product liability laws, and their content and interpretation varies even between relatively homogenous and partially harmonised EU product liability laws as the EU study on product liability conducted by John Meltzer, Rod Freeman and Siobhan Thomson in 2003 make explicit[864], the effectiveness of liability rules in steering behaviour is seriously hampered. When delivering software to a multitude of jurisdictions, few companies can afford to buy expertise from such a high number of lawyers as the ex ante analysis of the standard of care and the extent of liability would require.

*Impoverished sanctions.* The expressive effect of product liability rules is seriously hampered by the vagueness of the standard. But neither does the external constraining, which is attributable to the additional costs incurred by the sanctions attached to the breach of the product

---

[863] These problems are somewhat eased by the tendency to attach the consideration of the standard of care to cost-benefit analysis. Instead of negligence consideration being purely based on written norms and deriving its substance from the considerations of relevant laws and other regulations, court cases (especially precedents), instructions and recommendations given by officials, market practices and internal guides of firms etc., the standard of care is increasingly being established on the basis of some sort of a cost-benefit analysis. This is so even under the EC Product Liability Directive that explicitly establishes a consumer expectations approach to the analysis of the required safety level. As Thomas Wilhelmsson and Matti Rudanko note in *Tuotevastuu*, p. 158, with references to studies on UK and European product liability, any liability test short of absolute liability could be perceived as ultimately cost-benefit based. Under this approach, a well performed and documented cost-benefit analysis for each possible vulnerability and threat could inform the vendors of the appropriate behaviour. However, there is among the legal experts certain reluctance to accept such a cost-benefit approach that strives for economic efficiency and, at minimum it typically is combined with moral considerations. Of this reluctance in the Finnish tort law doctrine, see the analysis of Mika Viljanen in Ihmisen identiteetti ja tuottamusarviointi, p. 440-6.

[864] Meltzer et al., *Product Liability in the European Union.*

liability rules survive without serious impediments from the substance of the rules. The effects that the threat of legal sanctions can cause, depends in part upon their strength. As Hugh Collins explains in relation to the regulation of contracts, the private law system of regulation in general suffers from the structural weakness of impoverishment with respect to penalties for violations of the rules[865]. This seriously hampers the external influence mechanism that product liability rules have in theory.

> "Liability to pay damages in order to compensate for losses caused by breach of mandatory rules and contractual provisions imposes a cost on businesses for breach of the rules, and therefore provide an incentive for compliance. The strength of the incentive depends upon the measure of liability. Under the traditional perspective of private law, the measure of damages has almost invariably been restricted to one of restoration or compensation for provable losses. This corrective justice measure regards the purpose of the remedy as one of restoring the balance of the wealth of parties, either as considered prior to the formation of the contract or damaging action, or as it would have been after proper performance of the contract. This narrow policy objective excludes the possibility of setting the measure of liability at a level designed to achieve an optimum level of compliance with the regulatory standard." (Collins, *Regulating Contracts*, p. 90-91)

The threats offered by the private law system of liability to pay compensation for the costs of cure of defects appear minimal. The sanction imposed by a court in the form of a liability rule provides little incentive to manufacturers and retailers to reduce defects in quality and security. However, private law can produce interesting alternative sanctions to the compensatory liability rule. E.g., the law can enhance the non-legal sanction of rejection of goods by lowering the threshold of quality defects that permit rejection. A compulsory duty to repair faulty goods might also improve the incentives for compliance. An obligation might be placed upon the manufacturer

---

[865] Collins, *Regulating Contracts*, p. 90-91.

to repair or replace any defective goods for a period of time after purchase, with this obligation enforceable by a court order of compulsory performance backed up by the threat of deterrent sanctions such as fines.[866]

In some jurisdictions, penal regulation, particularly in relation to consumer contracts, provides a more substantial deterrent against misleading trade practices. Similarly, some jurisdictions such as in Finland, as discussed above, employ administrative agencies to monitor and deter unfair contractual practices committed against consumers. These types of measures seem to be effective on the whole in deterring the kinds of abuses at which they are directed no doubt because they address the twin problems of access to justice and the weakness of sanctions. They control behaviour in the market to a degree which the ordinary private law compensatory measures of damages awarded by a court cannot. However, they operate only at the fringes of markets, deterring overly abusive practices.[867]

This poverty of legal sanctions and their consequent disability to influence the decision-making of the software vendors vanishes when widening the perspective of the costs of legal liability from pure legal sanctions to cover also non-legal sanctions such as refusal to do business and damage to reputation. Legal doctrine recognises that product liability risk is often wider than mere damages or costs due to specific performances and litigation costs. The economic consequences of product recall and/or repair and the negative publicity together with possible shunning by buyers caused by

---

[866] Collins, *Regulating Contracts*, p. 299. Although private law has the capacity to alter its remedial sanctions, courts have been reluctant to depart from the tradition of compensatory damages, as Hugh Collins, *Regulating Contracts*, p. 92, demonstrates in relation to English law. This reluctance derives in part form tradition, in part from the conception of private law as a system of principles of corrective justice rather than one of regulation, but also in part form a more justifiable concern that damages in excess of compensation might produce overly cumbersome difficulties for conducting business in the form of overly harsh and unforeseeable liabilities, resulting bankruptcies and huge problems for the insurance system (i.e., the floodgate argument against the expansion of liability).

[867] Collins, *Regulating Contracts*, p. 118.

damaging products can be orders of magnitude larger than mere damages. These costs, such as the consequent loss in stock value[868], give firms an additional incentive to avoid products liability litigation. This provides another reason to believe that seller liability would increase software security.[869]

> Even though these additional costs resulting from the use of non-legal sanctions are not attributable to the liability rules as such in the form the normative restriction made in the analysis of the influence mechanisms required, they cannot be and should not be omitted altogether in a regulatory analysis. In fact, their influence ought to

---

[868] Mark Geistfeld in Products Liability, p. 362, refers to several studies examining the impact of prominent product liability lawsuits on stock prices from more traditional goods than software. They indicate that news stories reporting on products liability suits significantly decrease a firm's stock value and that safety-related administrative actions like product recalls substantially reduce stock prices. In all of the summarised studies, adverse publicity concerning product safety costs the firm more due to the reduced stock value than does the associated liability or recall costs. Similar findings have been presented also in relation to information system security of organisations by Cambell et al. in The Economic Cost of Publicly Announced Information Security Breaches in 2003. They provide evidence of significant negative stock market reaction for published information security breaches involving unauthorised access to confidential data, especially personal data and consumer electronics. Information security concerns have already led to product recalls in consumer electronics. In Japan 3rd generation mobile phones have been recalled due to a software flaw enabling third parties to see private information stored in the devices. According to Yurie Ito from Japanese Computer Emergency Response Team Coordination Center JPCERT/CC http://www.jpcert.or.jp/english/ [23.2.2006] the reason for the costly recall has been the foresight of even more expensive product liability litigation. Yurie Ito (2005) Vulnerability Trends in Consumer Electronics, unpublished presentation in Midnight Sun Vulnerability and Security Workshop Retreat, 21st of June 2005 in Hailuoto, Finland (on file of the author).

[869] Note that the non-legal sanctions are, as Hugh Collins emphasises in *Regulating Contracts*, p. 114-117, related to the possibilities of market actors themselves to solve some of the problems related to market failures. But even though these mechanisms (investment in brand names and other sunk costs, guarantees etc.) are not enough to solve the market failures, the non-legal sanctions (provision of negative info etc.) that target the reputation and trust sought by the self-corrective actions, create central incentives for the vendors.

> be considered before legal regulation is introduced. When they are considered, the role of law diminished in proportion. The direct costs of legal liability can be marginal in the considerations of producing secure software compared to these other costs[870]. This is so despite the fact that the costs of liability in the case of information goods like software can be extensive due to the defects existing in every copy and the liability thus concerning every user.

Currently, these costs have not been foreseeable enough to provide sufficient incentives for secure software development from the societal point of view. Patching provides a reasonably low costs alternative to product recall and repair as has been argued above. Customers have also tolerated the vulnerabilities due to several historical reasons and for their lack of capability and information to compare products in terms of security. Consequently, shunning has not been a real risk for software vendors, even though the climate is changing. Customer tolerance of insecurity might be vanishing over time. It remains to be seen whether this is sufficient to improve the security of software in a sufficient manner but, as the theoretical considerations presented above predict, it is not likely to do so to the level sufficient to the networked society as a whole.

---

[870] As Collins notes in *Regulating Contracts*, p. 174, legal sanctions for breach of contract are rarely as important as other informal sanctions such as damage to reputation. There is also preliminary empirical evidence (e.g., Irlenbusch, Relying on a man's word?, and the works cited therein) from experimental studies that indicate the role of social norms in contracting and that trust, together with non-legal sanctions, plays a more central role than legal sanctions. The legal process or any other institution that is respected by the trading community as an impartial and reliable finder of fact or truth (e.g., consumer ombudsman, committee of a trading association or an arbitrator) does, however, contribute to the imposition of informal sanction. They function as an organised reputation mechanism collecting and disseminating what participants in the market regard as reliable information about trustworthiness. This type of contribution can happen even without ever invoking the power to impose legal sanctions. The influence upon contractual behaviour stems instead from the independent reactions of participants in the market to the authoritative judgements about business reputation issued by these respected bodies. (Collins, *Regulating Contracts*, p. 123-125)

*Of duties to disclose.* These considerations of the substance of the wide legal basis for product liability apply, in addition to the security standard setting function, similarly also to the effectiveness of these rules in setting duties to disclose and influencing behaviour by informative means. However, there are certain specific duty-to-disclose issues that have to be considered separately.

Default rules on defectiveness (conformity) essentially are signs of the market-compliant character of sales-laws measures in the sense that they provide obligations to inform about the harmfulness of a product[871]. However, the duties to inform of safety critical issues are most forcefully visible in consumer protection rules for the protection of the weaker party, especially in the consumer product liability rules. Despite the increase in the duties to disclose under the doctrines of good faith all over Europe, the duties to disclose even safety information are still exceptions. The need for the protection of the informationally weaker party has to be justified separately; no direct generalisations of the informational market failures in the market economy can be made.

In consumer protection cases the weaker informational position is clear. However, also a general obligation to inform about security related vulnerabilities also in enterprise relations can be supported as a measure protecting the informationally weaker party. The normal legal presumptions of approximate equality of bargaining power and comparable sophistication in evaluating benefits and risks are simply not correct in the context of software security. A typically purchaser (even industrial), excluding only the largest industrial buyers that have the resources to use the expensive and scarce expertise, is only slightly more experienced than a consumer and should not thus be treated as having sufficient information for rational decision-making on costs of software vulnerabilities[872].

---

[871] Wilhelmsson, Contribution to a Green Sales Law, p. 282.

[872] Even though industrial buyers are in a better informational position to evaluate the products and vendors claims than consumers, they still are at the mercy of the vendor largely due to the lack the means (not only are white-box methods largely unavailable, but also the tools for testing even on the

Buyers' (including industrial buyers') possibilities to assess the vulnerability of the software and to conduct security testing are reduced. This is not just due to the access to source code issues, but also due to the diminishing expertise of software development among information system developers (increased use of components and outsourcing instead of in-house development). All but the largest and most capable buyer organizations lack the resources or expertise to evaluate the security of a software product. Even though the higher stakes involved for industrial buyers justify measures to avoid the common information-deficiency problems that consumers face, the outright unavailability of information on the quality aspects of security together with the high costs of acquiring it due to the lack of methods to gather such information, and the underdeveloped state of metrics in the quality aspect of security, means that not even industrial buyers have sufficient information for sound decision-making.

More importantly, no reasonable and knowledgeable person would expect them to be able to do so. Buyers that search for information only as much as it is justified on cost-benefit basis and especially in the case of consumers, where the information acquisition is even more constrained, cannot be expected to use scarce and expensive experts to gather information that the COTS software vendors could much more easily deal with[873].

Despite the general duties to disclose vulnerability information even in enterprise relations on the basis of liability rules, there are differences in the wide legal basis for products liability in their obligations for the vendor to find out the harmfulness of the product and to inform correspondingly. Product liability rules under the EC Directive (85/374/EEC) not only obligate to provide safety information that the vendor happens to posses when concluding the

---

developer side are relatively undeveloped and not in wide use) and financial resources to evaluate the products security (high costs involved).

[873] Exceptions are buyers with special needs for security, such as many mission-critical governmental systems. Many of the big established COTS vendors have also started to license their source code for security purposes for these customers.

contract, but also to disclose the information that the vendor ought to have known (i.e., not just what she knew). The vendor is expected to know the characteristics of her products and has an obligation to find out the quality and security of the marketed products[874]. The obligation to find out the risks involved with the use of the product (such as its vulnerability) and to inform correspondingly is less clear and weaker in general contractual and non-contractual liability rules.

The dispositive nature of general contract and tort law rules affect the incentives to provide information. Since the buyer cannot claim as a defect something she was informed about prior to completion of the contract, the vendor is incited to inform only of the sub-normal quality or security; i.e., if she does not inform, the buyer can expect "normal" quality and security from the software under the default conformity rules[875]. However, not even sub-normal quality or security has to be disclosed in order for the vendor to avoid liability under dispositive default contract law rules on the defectiveness of products. The vendor can merely state that she does not know whether the goods are of normal (generally expected quality or security as defined by the default rules on sales of goods) and if the customer accepts

---

[874] This obligation is most clearly visible in the state of the art –defence (development risk defence) offered for the vendors as a way of escaping liability (art. 7(e) of the Product Liability Directive 85/374/EEC). The vendor thus has to prove that she under the state of current (at the time of releasing the product) scientific and technical knowledge was not able to discover the defect. The degree to which the vendors have to find out the harmfulness of the product can thus be relatively high in cases where the public at large can be considered to expect high level of safety. Note that in Finland and Luxembourg, no such defence is even offered for the vendor, and the obligation to find out the harmfulness of the product is even higher, i.e., product liability is also for those defects that the vendor could not anticipate or take into consideration at the time of releasing the product.

[875] This can be sufficient for the whole industry to start providing complete information as the basic hypotheses of economic theory predicts. When at least those above the minimum quality or security disclose, others in the market also have an incentive to disclose since consumers will rationally assume that critical piece of information such as its quality will be omitted only if the value of that attribute for that product is low. Note that the many informational market failures discussed hinder this as has been argued above.

this, then the agreed lower standard is applied[876]. Thus, no incentive to acquire information is created; only incentives to disclose the information she actually or customarily possesses are created[877]. In the EU law, this applies even to the sale of consumer goods[878].

The situation is different when, like in the Nordic contract law doctrine, the general disclaims of application of default defectiveness (conformity) rules without expressive contractual specifications leads to the application of the "as is" rules present both in commercial and consumer contracts[879]. According to these rules if the goods have been sold subject to an "as is" clause or similar general reservation on their quality, the goods shall, nevertheless, be considered defective if the goods are in essentially poorer condition than the buyer reasonably

---

[876] As Steven Matthews and Andrew Postlewaite in Quality Testing and Disclosure, p. 334, note with references to similar results made by others, the announcements of ignorance can be credible if testing or disclosing is costly; in such cases the vendors are allowed to continue not the test and not to inform without fear of negative inference by the customers. In the software security this can be the case due to the high costs involved in the testing for security.

[877] Note that this can be socially suboptimal in terms of injurers incentives to obtain information about risks that she may have engendered. See Shavell, Liability and the Incentive to Obtain Information about Risk, p. 268.

[878] For example, as Riesenhuber argues in Party Autonomy and Information in the Sales Directive, p. 353, the Sale of Consumer Goods Directive 1999/44/EC (Directive 1999/44/EC of the European Parliament and of the Council of 25 May 1999 on certain aspects of the sale of consumer goods and associated guarantees, OJ L 171, 7.7.1999, pp. 12-16) does not impose an obligation on the seller to examine the goods or have them examined by competent person before the sale. The seller merely has to tell the consumer what she knows and what she does not know about the goods. No obligation to inform about risks that could not be anticipated when releasing the product considering the state of the art (development risks) is given. Not even risks that result from the sub-normal safety (is not designed according to the general safety knowledge of the branch) has to be disclosed; the vendor can merely state that she does not know whether the goods are of normal quality and safety and if the buyer accepts this, then this lower standard is accepted even in EC sales of consumer goods. These default rules are similar to the general dispositive contract law rules on defective products in national laws.

[879] Finnish Sale of Goods Act 355/1987 section 19; Finnish Consumer Protection Act 38/1978 chapter 5 section 14.

could expect[880]. There is even no requirement of essentiality in the considerations of the poorer quality in consumer contracts. Thus, in order for the liability for defectiveness to be limited from this "as is" level, the accepted defects must be explicitly stated[881]. No general disclaimer is thus valid and the vendor is incited to acquire information on the defects to the degree customers can reasonably expect[882].

Despite of these vagueness's in the information gathering and disclosure inciting provision prior to concluding a contract, all of the above mentioned rules create duties to inform about the safety of the product when new relevant information is gained *post-sale*[883]. Thus, the seller has to try to inform the users of the changed harmfulness of the product, e.g. when someone brings the existence of a vulnerability to the knowledge of the vendor. However, whether or not such duty exists when research provides new insights that becomes known among the professionals in the field depends on whether the product liability rules, indispositive consumer contract law rules, or

---

[880] Note that the seller's liability for defectiveness in cases where she failed to disclose to the buyer facts relating to the properties or the use of goods is explicitly stated in these provisions (subsections 2 point 2). Additional conditions for the liability are that the seller could not have been unaware of the facts (can be expected to know), that the buyer could reasonably expect to be informed about them, and that the failure to disclose can be presumed to have had an effect on the contract. These conditions are currently fulfilled at least in relation to the widely known security related vulnerabilities (like buffer overflows).

[881] Hemmo, Vuoden 2000 ongelma ja siviilioikeus [The year 2000 problem and civil law], p. 30-31, footnote 46. This is generally accepted at least in the Finnish and Swedish doctrines as noted by Mika Hemmo in *Sopimusoikeus II*, p. 123.

[882] Since the duty to warn developed in legal practice in Finland covers not only the information the vendors know, but also what they should have known, the incentives for vulnerability information acquisitions are optimal in the form explicated by Steven Shavell in Liability and the Incentive to Obtain Information about Risk, p. 265 and 268. They provide incentives only in those cases where the social value of information exceeds the costs of information.

[883] Note that information that is provided prior to purchase is more effective since it allows users to compare products before any lock-in, due to switching costs, happens.

dispositive (commercial or consumer) contract law rules apply. There are also differences in national doctrines.

*Contractual practice.* The default defectiveness rules in contract and product liability together with marketing rules and tort law rules seem to provide incentives both for developing more secure software and disclosing vulnerability information to the customers. There are limitations and vagueness's but by making the assumption that these rules are applied and effective they seem to provide effects that are overall beneficial for secure software development.

However, as already emphasised above, the substance of product liability especially in the software context is formed in the contracts and contracting practices of the software business[884]. Contracting parties, especially outside specific consumer product liability regulations such as the EC Product Liability Directive (85/374/ EEC) and national consumer protection acts, are free to agree differently from the default liability rules established in contract law[885]. Due to the dispositive nature of majority of the product liability rules especially in enterprise relationships, the substance is primarily determined in contracts. The main norm source for the quality and security levels that the buyer and user of software has a right to expect

---

[884] Note that research on the contractual practices and negotiation strategies in security related issues is scarce. This is why these considerations can be only tentative at best. The lack of research concerns the contractual practices and use of contract terms in the software business in general. This could provide an interesting research area to be taken further.

[885] In non-contractual (delictual) relations such disposition are not possible due to the lack of cooperation part of which the liability could be agreed upon differently from the default allocations. However, warnings and information provided about the safety of the product and about the extent of the liability of the manufacturer can function like contractual liability limitation clauses. For example, information given about the security of the product and its proper use can be considered to alter the risk awareness of the user and thus the care required from her. If the user behaves contrary to the given information, tort liability can be limited on the basis of contributory negligence or neglect of the duty to limit the resulting losses. Of this in relation to Finnish law see Hemmo, *Sopimus ja delikti*, p. 188-191, and Mononen, *Yritysten välinen tuotevastuu*, p. 173-174.

stem from the contracts and complementary marketing information. The business practice is a secondary source, and the default rules in sales of contracts regulations apply only if these legal sources are silent or unclear[886].

The problem is that contracts for COTS software are typically silent about the security level to which the software should conform to and have for long lacked specific clauses concerning quality requirements[887]. Such clauses are not present in typical license agreements and standard form contracts that emphasise the disclaiming of all liabilities or at least exclusion of most of them. Especially in consumer contracts such provisions are close to non-existent and still rare even in commercial contracts for COTS software. Neither is information on the (quality aspect of) security of the software typically given in the marketing or other pre-purchase information[888]. This is so despite the rising demand for secure software.

---

[886] Above, reference was made to the agreement of a lower quality standard than provided by the default rules in the sales of goods laws. In here, concentration is on a higher agreed standard which as such provides information of the quality and security of the product that the buyer can expect. The vendor is liable if she does not conform to this higher agreed standard. Such dispositions of the higher standards are possible also under the EC Product Liability Directive (85/374/EEC) even though it prohibits dispositions about the liability which is lower than what the directive provides. See article 12 of the Council Directive 85/374/EEC of 25 July 1985 on the approximation of laws, regulations and administrative provision in the Member States concerning liability for defective products, OJ L 210, 7.8.1985 pp. 29-33.

[887] Note that COTS software typically does not include a detailed specification, if a written specification of requirements even exists. This is why it may be difficult to judge whether a bug is in fact a defect entitling the customer to damages solely by reference to the specification.

[888] According to the defect (conformity) rules of Nordic Sale of Goods Acts (section 18) and similar provision in consumer protection legislation, a product is defective if it does not correspond with particulars given during its marketing. Similar provision are present also in the article 2, paragraph 2(a,d) of the Directive 1999/44/EC of the European Parliament and of the Council of 25 May 1999 on certain aspects of the sale of consumer goods and associated guarantees, OJ L 171, 7.7.1999, pp. 12-16 (Sale of Consumer Goods Directive).

For long, customers did not see the need to negotiate such terms or expect marketing information to raise security issues. Currently, other than sufficiently big influential actors such as governments and large international companies do not have the leverage to negotiate such terms into the standard contracts separately. Even in such contracts the security related vulnerabilities as credence characteristics (i.e. they cannot be appropriately assessed even after consumption or only a long time after that when, e.g., an attack occurs) are likely to be such defects that fall outside the specifications.

The disposition of the quality and security demands of the products as such must be separated from the disposition of the liability rules. In the former case the purpose is to alter the standard to which the software must conform to. In the liability disclaimers and limiting clauses the purpose is to shift liability to the contracting party (in the COTS software context typically to the buyer or user) and to alter the extent of the liability from that what it could be under the default rules.

The degree to which liability issues are contracted differently affects the substance of the product liability rules and changes their regulatory effects. Here we come back to the above mentioned starting-point that the default defectiveness rules apply also to COTS software vendors.

The extensive uncertainty of the applicability of liability rules and especially their possible extent has led software vendors to use restrictive liability clauses and disclaimers. Agreeing on liability differently from the default rules provided especially in contract law is a central part of contracting particularly in the COTS software business. The liability disclaimers and exclusion clauses typical in the software business at minimum restrict liability for consequential damages (indirect losses) that result from losses of income or production, losses of processed data etc. and often also limit the liability to the price of the software or certain ratio bound to it in all circumstances[889].

---

[889] This is the case for example, in the multilaterally drafted Finnish contract template IT2000 Terms and Conditions for IT Procurement, YSE General Terms and Conditions, where section 9 disclaims all liability for any indirect or

It is a form of risk management for the vendors. The strategy of software vendors is to avoid contractual, tortious or specific products liability by shifting the burden set by default rules to their customers or to the injured parties. By disclaiming liability the risk of liability for software defects is managed relatively easily and with low costs in an international context with large customer base.

The use of clauses limiting the liability of the vendor has been justified by the peculiarity of the line of business[890]. The COTS software business has been considered to have such peculiar characteristics that it needs special rules of liability.

> The reasons for the almost universal use of contractual liability disclaimers are diverse. First of all, software producers may be exposed to a greater degree of liability risk than the producers of more traditional goods. A defect in a software product is likely to exist in every copy of the software made, dissimilar to traditional industries where a defect done during manufacture is likely to exist only in some of the products made. This increases the possible liability costs of software vendors since they have to answer to every user instead of just few.

---

consequential damages, and limits the liability for direct damages to 15 percent of the price of the software. Similar rules are applied to the sale, licensing or other assignment of rights of use of standard software products under the section 10 of the IT2000 PTE Terms and Conditions for Small Scale deliveries. The latter types of disclaimers are especially beneficial for the vendors since the liability risk cannot expand uncontrollably and the confines of the liability are known right after the contract has been made. However, the practice of limiting liability extensively is not special to the software business. As Hugh Collins, *Regulating Contracts*, p. 228, notes in relation to the standard form contracts in general: "A closer inspection of the terms of the standard form contracts usually reveals a pattern in which the risks are allocated routinely to the buyer by extensive use of disclaimers and exclusion clauses. The effect of such clauses is that the business retains a discretion under the contract over the level of quality and security which it will provide, whereas the buyer is bound to make payment, expect perhaps in the event of complete failure of performance by the business".

[890] The contractual characteristics of COTS software business have been empirically studied by Juhani Warsta in 2001, *Contracting in Software Business*. See especially page 182 where these characteristics are gathered in a table format.

The use of liability disclaimers in COTS software business is further justified by the fact that there are numerous customers and the relationship with the customer is weak[891]. This means that the vendor has little means to evaluate the potential risks especially from consequential damages due to the variety of unpredictable ways the product may be used, and the capabilities to avert or limit and insure against such damages in any cost-efficient manner lies at the customer[892]. However, this argument seems unconvincing as noted by Hugh Collins[893]. Since the default terms provided only apply to ordinary uses of the goods, unforeseeable losses are not therefore recoverable under the liability rule on the ground of remoteness. To exclude the default terms on satisfactory quality on these grounds is therefore to protect the vendor against non-existent liability. It is especially for such reasons mandatory regulation, that sets the quality provisions as compulsory terms of the contract, requiring the supply of goods of satisfactory quality in consumer standard form contracts, has been imposed.

Also the uncertainty of the extent, and difficulties in quantification and insurance of the typical financial losses resulting from a vulnerability increase the vendors' fear of exposure to crippling legal actions. Especially in the case of distribution of software through the Internet, where relatively inexpensive products are offered, the liability disclaimers have been considered to be especially important

---

[891] As Warsta (2001, p. 134, 138 and 212) shows in analysing the contract process and relationships in the software business, customers (end-users or system developers) may stay totally unknown to the software company especially when the application is procured either from the retailer or directly downloaded from the Internet. The software can be just bought and installed for use without further cooperation needs. This is the case in other than delivery channel operations, where relationship especially with the distributors is closer.

[892] Even if the vendor could estimate the risk of such losses, it would have to charge higher prices from all customers to cover the risk of unusual consequential losses. The better possibilities of the customers or other injured parties to limit the risk of consequential damages by setting up firewalls and using virus-scanners etc. (bilateral precaution) is an argument favouring the acceptability of contractual provisions excluding liability for consequential damages in the Finnish doctrine (Hemmo, Vuoden 2000 ongelma ja siviilioikeus [The year 2000 problem and civil law], p. 18-19).

[893] Collins, *Regulating Contracts*, p. 291.

> and are typically limited to the price of the software or certain ratio bound to it in all circumstances[894].
>
> The main theme behind these arguments, which is at the same time one of the most compelling justifications for the use of extensive liability disclaimers and exclusion clauses, is that they enable innovation in young markets. Fear of crippling legal actions is considered to cause vendors not to sell new products that could prove to be widely useful but also imply significant risks. When new innovative firms can efficiently minimise the liability risk, they are more likely to bring new products to the market quicker. Customers that are willing to take the risk of using possibly unfinished products can do so and take the possibly highly useful piece of software into use earlier.[895]

The liability standards that deviate from the default allocation provided by law are sometimes drafted separately by the contracting parties for each transaction, but in the case of COTS software they have typically been codified into standard contract templates in order to facilitate the contracting procedure and to save transaction costs.

---

[894] Warsta, *Contracting in Software Business*, p. 187. Note that the price of the software provided is not just a justification for the use of the limitations of liability in general, i.e. that risks of relatively cheap but business critical products causing wide damages are too big for many companies to take. It can also be an indication for the buyer of the expectations that she can place on the coverage of liability. This can be seen to limit the liability of the vendor as such, without explicit liability disclaimers as Mika Hemmo, Vuoden 2000 ongelma ja siviilioikeus [The year 2000 problem and civil law], p. 47, points out in relation to the Finnish legal system in 1999. The attachment of the expectation of the buyer partly with the price of the software (i.e., the quality you can expect depends partly on the price that you pay for the software) also mitigates the concern expressed about the application of the same liability rules to the volunteer contributors of open source software who do not get compensated and do not receive a stream of income with which to purchase liability insurance. When no fee is required in any form from the use of open source software, the potential liability of the developers of open source software is much more limited than the liability of the COTS software vendors. What this means is that increased liability for security flaws may improve the security of COTS software, but it may not improve open source security as pointed out by Fred B. Schneider in Open Source in Security, p. 126-127.

[895] We will come back to this liability and innovation discussion in the section concerning the reactions of objects.

It is these fixed software contract templates, characteristic for software business, which define in most detailed way the substance of product liability rules[896].

The enforceability of these liability disclaimers and exclusion clauses in shrink- or click-wrap or similar types of standard form contracts typically used in the software business is a still matter of dispute[897]. As the study of Andrea Migdal from 1999 about the enforceability of shrinkwrap licenses in several jurisdictions show, jurisdictions have differed substantially on these matters; some expressly have forbidden liability disclaimers in standard form contracts, others have generally enforced such licenses, and some with limitations[898]. Similarity exist only in that typically the disclaiming of liability is accepted in courts only to the degree the breach of contract or duty is not intentional or consequent to gross negligence[899].

---

[896] This is no specialty of the software business. As Collins, *Regulating Contracts*, p. 63, explains in relation to regulation of contracting: "Most detailed standards governing a particular transaction are set by the parties themselves and described as the terms of the contract. In some instances these standards will be the result of negotiation, but more typically they will be set by one of the parties and contained in a standard form contract".

[897] This has recently been pointed out also by two practitioners of UK law, Dominic Callaghan and Carol O'Sullivan in their article on the liability for software errors from the beginning of year 2005, Who Should Bear the Cost of Software Bugs?, p. 57, in relation to the legal systems of most EU countries.

[898] Migdal, Shrinkwrap Licenses Abroad, p. 751-753. The different approaches among the legal systems of EU Member States to the validity of liability exclusion clauses in general has been pointed out by Christian von Bar and Ulrich Drobning, *Study on Property Law and Non-contractual Liability Law as they relate to Contract Law*, p. 177-178, in their study on the problems and obstacles for the smooth running of the internal market resulting from differences in property law, non-contractual (tortious) liability law and contract law in most of the legal systems of the European Union. The approaches differ both in the validity of the exclusion of contractual liability and the exclusion of non-contractual liability.

[899] This principle is repeated in basic study books about contractual liability. For example, Mika Hemmo notes in *Vahingonkorvausoikeuden oppikirja*, p. 206-207, the internationality of the principle that the party who has a duty to perform under a contract cannot receive protection from liability disclaimers and exclusion clauses from the liabilities resulting from grossly negligent or

Internationally the development has for long been leading to wider acceptance of the standard form terms as part of the contractual agreement also in the software business, i.e., that the formal requirements for the inclusion of the standard terms are fulfilled[900]. As Jennifer Chandler points out in relation to U.S. and Canadian law in 2005, "[t]he software purchaser is likely to be bound by these contractual terms [authors note: disclaiming warranties and conditions, and clauses either limiting or excluding liability for consequential damages in end-user license agreements EULAs], and may face a significant hurdle in convincing a court to set them aside"[901]. Similar development has been reported also from the United Kingdom that has been considered one of the European countries in which the supplier has been least likely to be able to enforce any clause which limits or excludes liability[902].

---

intentionally damaging behaviour.

[900] This has been pointed out by Trompenaars and Hugenholtz in *Formation and validity of On-Line Contracts*, p. 45, with reference mainly to U.S. legal literature on online-contracts. Such formal requirements include the users awareness of the existence of the license and that the transaction is subject to it, the users possibility at minimum to read the license terms before ordering or executing any download commands and, especially in relation to online licenses, that the user takes affirmative steps to signify agreement, e.g., by hitting "I Accept" button or similar.

[901] Chandler, *Improving Software Security*, p. 10.

[902] Dominic Callaghan and Carol O'Sullivan in their article on the liability for software errors from the beginning of year 2005, Who Should Bear the Cost of Software Bugs?, p. 58, point out that at least in IT contracts between two commercial parties, the Courts in the U.K. have been more reluctant to invalidate clauses limiting or excluding liability. The Courts' view is that such commercial parties will have negotiated a price which reflects the allocation of risk that has been achieved by the exclusion or capping of liability. Note that in the case of allocation of security risk this might not be the case primarily due to the inadequacy of information about the quality aspects of security. For the reasons presented in section 2.8 above, all but the largest and most capable buyer organisations lack the resources or expertise to evaluate the security claims made of a software product. We will come back to this. Callaghan and O'Sullivan concentrate on the quality aspects and only briefly refer to the security aspects of software bugs; they do not deal with the special issues of secure software development and especially not the quality aspects of security.

This is the case also in Finland. However, the legal doctrine in Finland has reacted to the wider validity of standard form contracts in general in order to protect the weaker party. One such reaction is the emphasis of content-based considerations in the conditions according to which the standard terms become part of the contract[903]. In this approach towards the binding force of standard form contract terms, comprehensive considerations taking into account the characteristic features and requirements of the type of contract and the economic activity in question are used instead of formal criteria such as the manner or form in which the terms are presented[904]. This content-based approach is visible especially in the doctrine of surprising and severe standard terms according to which terms that significantly increase one party's obligations in comparison to default rules in indispositive law or to what is normal and ordinary for the type of contracts in question have to fulfil more stringent criteria in order to become binding[905].

---

[903] Another development is that the unclear standard terms are interpreted to the detriment of the party who drafted them. Also perfectly valid liability disclaimers and exclusions clauses in standard form contracts are interpreted narrowly, i.e., in a way that limits the liability as little as possible. (Hemmo, *Vahingonkorvausoikeuden oppikirja*, p. 206; Hemmo, *Sopimusoikeus II*, p. 75; Wilhelmsson, *Vakiosopimus*, p. 55 and thenceforth)

[904] See, e.g., Wilhelmsson, *Vakiosopimus*, p. 88 and thenceforth.

[905] The idea behind such "semi-compulsoriness" of dispositive default contractual rules is that when the legislator or other regulator in the case of contractual rules (e.g., a court or development through contractual practices in the market in question) has after thorough preparation enacted statutory rules to govern certain activity, those rules include an evaluation of what constitutes equal treatment of the parties in the relations in question. Deviating form such default allocations that may have significant policy choices behind them by contractual means is only possible with good justifications. One-sided benefit at the other party's expense is not held to constitute appropriate grounds. This established approach is explained even in the basic introductory texts on Finnish law like the one written by Juha Pöyhönen in English, The Law of Obligations, p. 91-92. See also the recent basic textbook on Finnish contract law written by Mika Hemmo, *Sopimusoikeus II*, p. 293. Jan Hellner in *Lagstiftning inom förmögenhetsrätten*, p. 160 explains this approach in relation to the Swedish contract law.

Even though there is a rough consensus that the clauses disclaiming or excluding liability in software licenses are valid in business relations due to the reliance on the principle of freedom of contract in most western countries, their validity in relation to private consumers is a heavily disputed and controversial issue. A dominant pattern in the European regulation of standard form contracts issued to *consumers* is to employ provisions that permit judges to invalidate unfair, unreasonable, or unconscionable terms[906]. This means that if a dispute with a consumer reaches court, there exists a strong chance that the fierce exclusion clauses will be declared invalid and not binding on the consumer. The license terms in shrink- or click-wrap or similar types of standard form contracts might not be enforced in courts in relation to private consumers even if they are considered to become parts of the contracts in a correct way[907].

In certain jurisdictions, such as the Finnish, this applies even to enterprise relations[908]. Whereas the Unfair Terms in Consumer Contracts Directive (93/13/EEC) is limited only to standard form contracts (article 3 of the directive limits application only to "contractual terms which has not been individually negotiated") and applies only to contracts concluded between a consumer and an

---

[906] Main provisions can be found from the EC Directive 93/13/EEC of April 1993 on unfair terms in consumer contracts, Official Journal L 095, 21.4.1993, p. 29-34.

[907] This requires that the court take an explicit consumer protection approach. Hugh Collins in *Regulating Contracts*, p. 35, sets out the influence of doctrinal coherence as an obstacle in the way of occasional judicial attempts to help consumers by deleting unfair exclusion clauses. If the courts favour free markets and the classic principle of the freedom of contract according to which parties should be free to select the terms of their transactions over consumer protection considerations, courts would reject the possibility of striking down exclusions clauses on the ground of unfairness.

[908] The wider applicability of provisions that permit judges to invalidate unfair, unreasonable or unconscionable terms is noted in relation to software also by Elizabeth MacDonald in Y2K and Contractual Exemption Clauses published in 1999 in a special feature of the Journal of Information, Law and Technology concentrating on the Year 2000 problem.

enterprise, in Finland there are laws that apply both to individually negotiated terms and to enterprise relationships[909].

Despite of the doubtfulness of the validity of these terms in standard form contracts especially in consumer relations in many countries, they are widely used especially in the software business. Even though the validity of extensive liability disclaimers and exclusions clauses in standard software licenses, especially in end-use license agreements (EULAs), depends on the legal system and varies between them with differing lines of argument and emphasis used, the standard software contract templates disclaiming and excluding liability are used as if they were universally valid contracts. The highly complex array of different legal standards that set varying requirements for the validity of liability disclaimers and exclusion clauses and the unpleasant surprises when the national requirements are not satisfied, are overlooked. The substantial costs in ascertaining what the legal position is, even if it could be ascertained with anything like certainty, certainly are one explanation for this practice. But it is not a sufficient explanation and certainly does not hint anything on why the

---

[909] Provision on the regulation of contract terms both in the Finnish Consumer Protection Act (38/1978) chapter 3 and Act on the Regulation of Contract Terms Between Enterprises (1062/1993) [L elinkeinonharjoittajien välisten sopimusehtojen sääntelystä] make it possible for the Market Court or even for the Consumer Ombudsman to forbid enterprises, accompanied with a threat of a fine, to use contract terms that are considered unfair. Even though the injunction issued by Market Court does not directly influence the validity of the unfair terms already in use by the enterprise, neither do the use of such terms in the future become invalid as such (i.e., the breach of the injunction leads only to the enforcement of the conditional fine attached to it), it creates a strong cause for adjusting the forbidden term(s) in individual disputes between contracting parties. Adjustment of unreasonable contract terms is possible in individual disputes both in the relationships between a consumer and an enterprise under Consumer Protection Act (38/1978) chapter 4 section 1 and in enterprise relationships under general Contracts Act (228/1929) section 36 §. These provisions on adjustment of contract terms allow general courts of first instance and the Consumer Complaints Board in individual cases initiated by one of the contracting parties to adjust, to interpret or even to disregard unreasonable terms

disclaimers and exclusions clauses are not contested in the software business.[910]

The reason for this practice, which might give negative signals to the users about the quality and security of the software[911], becomes understandable when realising, as Hugh Collins explains, that the real force of these contractual clauses arises in the event of dispute[912]. By the combination of liability exclusion clauses and restrictions on possible avenues for legal redress (e.g., choice of jurisdiction clauses or arbitration clauses), the standard form grants the vendor a strong bargaining power in post-breach negotiations. When a complaint is made, the vendor can point to clauses that exculpate or greatly diminish the responsibility of the business, which serve to discourage

---

[910] It is interesting to note that similar liability disclaimers and exclusion clauses can be found both from proprietary and Open Source licenses as noted also by Miko Välimäki in *The Rise of Open Source Licensing*, p. 170. The inclusion of such clauses appears to be standard practice in both despite the questionability of the validity of these clauses. The use of potentially unfair terms in standard-form contracts is not, however, specific only for the software industry. Their ubiquity has been demonstrated, as emphasised by European Commission in its report on the implementation of Unfair Terms in Consumer Contracts Directive (COM(2000)248 final, p. 9), in several studies conducted by the Commission analysing certain types of standard-term contracts proposed to consumers in different Member States (e.g., contracts of sale, car rental contracts, insurance contracts, contracts in the field of air transport, contracts concerning certain banking services, tourist services and the provision of general interest services). For a short overview of these studies see Annex II of Commission report (COM(2000)248 final).

[911] By stating in their most strict form that the vendor is not prepared to take any responsibility for her product, the restrictive license terms might give a negative signal to the potential buyer about the quality and security of the software in question and could, if the customers become serious enough for security, harm the sales of the software. However, as long as the liability disclaimers are effectively explained in terms of the special features of the software business and consequent risk allocation, the signal is not considered negative. Only in situations where these explanations would not be considered valid, could the signal possibly be considered negative. This remains to be seen when, and if, the liability disclaimers become seriously questioned on security grounds either via widely published court practices or other public discussion.

[912] Collins, *Regulating Contracts*, p. 229.

the pursuit of the complaint[913]. Even if the terms are of doubtful validity, there is in fact good reason for leaving them in as a source of bargaining strength in post-breach negotiations. Since the consumer is most unlikely to litigate the issue before a court, as will be discussed later on in the analysis of the implementation, the objectionable clauses will never be formally invalidated, and so they can be relied upon at least for the purpose of negotiating or imposing a settlement on the consumer.[914]

This argument about the post-breach negotiation bargaining power increasing function of the liability exclusion clauses in standard form contracts emphasised by Hugh Collins[915] also explains why the rare existing invalidations of the terms in standard form contracts by courts in individual disputes between parties have not been, and even if they would increase, are unlikely to be effective against the use oppressive license terms in software contracts. Their use continues even after they have been considered by courts as void. As Collins explains, the type of regulation which permits judges to invalidate unfair contract

---

[913] Contract terms stipulating that the compensation in any case is limited to the price of the software, for example, can in the case of COTS software being used as a component in an information system be an effective psychological barrier for claiming any damages.

[914] Maybe the argument made by Yates in his survey study about the use of exclusion clauses in contracts in the U.K in 1974-1976 [Yates D (1982) *Exclusion Clauses in Contracts*, 2nd ed., Sweet & Maxwell, London, p. 29. The study is reported in Collins, *Regulating Contracts*, p. 229)], about the reasons for companies lack of sanguineous to enforce exclusion clauses in consumer related mass contracts and about their use in situations where they are likely to be void and possibly even subject to minor criminal offences could be generalised similarly to Collins in *Regulating Contracts*, p. 229. As Yates explains "[w]hile they [exclusion clauses] would, as a matter of policy, always be relied upon, their insertion seemed to be largely a question of psychology" [quoted in Collins, *Regulating Contracts*, p. 229]. I.e., they leave the customer with the possible erroneous view that she had few enforceable contractual rights against the vendor or no feasible legal avenues to demand them. This psychological effect can even prevent the consumer from expressing her dissatisfaction with the product to the vendor and demanding compensation, not to even speak about taking the issue to court.

[915] Collins, *Regulating Contracts*, p. 233.

terms is unlikely to have any impact on the use and content of standard form contracts[916]. Since the use of oppressive terms in standard form contracts is not unlawful in itself, there is no reason to leave them out.

Even if a court determines that the clause is invalid, there is no reason to refrain from using the clause as a bargaining counter in contracts with other consumers. These customers are most unlikely to be able to discover whether or not a court has declared a clause invalid, and the precedent may always be distinguished by introducing a minor alteration in the contract[917]. Even when a list of terms that will invariably or normally be regarded as being invalid is provided in legislation (as in the annex to the EC Directive 93/13/EEC on unfair terms on consumer contracts[918]), it cannot provide a complete guide as to the sorts of terms which courts will find objectionable.[919]

---

[916] Collins, *Regulating Contracts*, p. 233. The somewhat sad conclusion of the Commission Report on the Implementation of Directive on Unfair Terms in Consumer Contracts (93/13/EEC) was that "[d]espite the legal mechanisms created to encourage the elimination of unfair terms in consumer contracts, such terms continue to be used on a wide scale" (COM(2000)248 final, p. 35).

[917] This problem is due to the results of the enforcement of a decision enjoining the elimination of an unfair term under Unfair Terms in Consumer Contracts Directive (93/13/EEC) being limited to the actual wording of the term itself. The effects of the term, i.e., the imbalance it creates between the enterprise and the consumer, that underlies the court's decision, lie outside the scope of the enforcement. This contradiction between the goal of the legislation on unfair terms and the results of its enforcement is emphasised in Commission report on the implementation of Directive on Unfair Terms in Consumer Contracts (COM(2000)248 final, p. 22-23). The possibility to deviate from default allocations of rules and to circumvent the possible invalidations of the rules in standard form contracts explains why the dispositive rules that do not conform to the real market practices and widely used mechanism have a weak influence on behaviour. Indispositive rules are not strong tools to change established market practices. This has been explained, e.g., by Jan Hellner in *Lagstiftning inom förmögenhetsrätten*, p. 162, who at the same time also acknowledges the direct behaviour influencing mechanism of dispositive contract law rules like those concerning defective products.

[918] Council Directive 93/13/EEC of 5 April 1993 on unfair terms on consumer contracts, Official Journal L 095, 21.04.1993, p. 29-34.

[919] Collins, *Regulating Contracts*, p. 233-234.

Under mandatory product liability rules like EC Product Liability Directive (85/374/EEC) and those in national sale of consumer goods laws, contract terms that differ from the standards provided in them to the detriment of the buyer or the injured person done prior to the damage are void[920]. This means that the limitation or exclusion of liability is not allowed with contractual means at all.

For example, the indispositive rules in the Finnish Consumer Protection Act 1978/38 section 20, according to which the buyer is entitled to compensation for indirect losses if the defect or loss is due to negligence attributable to the seller or if, at the conclusion of the contract, the goods differed from an express representation of the seller, cannot be contracted to the detriment of the buyer[921]. The lawful rights of the consumer under mandatory provision cannot be limited. This has been restated by the Finnish Supreme Court in a recent precedent 2004:123 concerning an injunction reinforced with a threat of fine sought by the Consumer Ombudsman from the Market

---

[920] According to the article 12 of the EC Product Liability Directive (85/374/EEC) "[t]he liability of the producer arising from this Directive may not, in relation to the injured person, be limited or excluded by a provision limiting his liability or exempting him from liability." Also the Finnish Consumer Protection Act (38/1978) chapter 5 on sale of consumer goods, as an example, states in section 2 that "[a] contract term differing from the provisions of this chapter to the detriment of the buyer shall be void unless otherwise provided below." It is important to note the restriction on the scope of these rules. They primarily apply only to the consumer who either has bought the product or suffers damages from its use. If the software is delivered to the consumer through a channel of operators (importer, retailer etc.), the mandatory rules do not make the liability disclaimers and exclusion clauses between the delivery channel operators void. Also other product liability than that basing on such mandatory rules can be agreed upon differently, as already pointed out above.

[921] Section 20, subsection 1, of the Consumer Protection Act 1978/38 reads as follows: "The buyer shall be entitled to compensation for loss that he/she suffers because of a defect in the goods. Indirect loss, referred to above in section 10(3) and (4) shall, however, be compensated by the seller only if the defect or loss is due to negligence attributable to it or if, at the conclusion of the contract, the goods differed from an express representation of the seller."

Court on the use of a contract term in a warranty for mobile phones that limits the seller's liability for indirect losses[922].

Such unambiguous statements about the impossibility to limit liability basically provide consumers with the knowledge about their rights to compensation and have a greater potential to shield from the psychological effect of the liability disclaimers and exclusions clauses in software licenses. Such indispositive rules on defective and insecure products hold a stronger expressive influence than the rules that merely allow courts in individual disputes to adjust the terms.

Still their informative role is limited and their real potential in shielding from psychological effect of the liability disclaimers and exclusions clauses can be doubted. For example, the contract terms stipulating that the consumer is not entitled to compensation for indirect losses arising in the case of a defective product may give an erroneous notion of the lawful rights of the consumer even under the mandatory rules. Consumers cannot be expected to be generally aware of the content of the mandatory provision on right to compensation for defective products in consumer protection laws. This is also the argument behind the Finnish Supreme Court ruling 2004:123 where the court, without taking an actual position, noted that a contract term stipulating that the company does not *in any case* have liability for indirect damages or losses might be deemed unfair due to the erroneous notion of lawful rights that such a vague formulation may give to the consumers[923].

---

[922] The case is available in Finnish at the Finlex Data Bank (http://www.finlex.fi). The Supreme Court ruled that contract clauses limiting all liability for consequential or indirect damages or losses cannot be deemed unfair from the point of view of the consumer when the terms make explicit, when judged in their wholeness, that they do not restrict the liability provided in mandatory legal provisions.

[923] The reason why the Supreme Court did not take an explicit position on the question whether the exact contract term used by the company in this case is unfair or not is that the Consumer Ombudsman had not invoked this basis for unfairness. Note that such obiter dictum comments do not have the status of a precedent and that under the Finnish legal system not even a judicial precedent is binding. The lower courts (Courts of Appeal and district courts) may depart from earlier decisions made by the Supreme Court and it can even itself depart

The use of software licenses gives the vendor an argument against the applicability of such indispositive rules, which can be used in post-breach bargaining similarly to the liability disclaimers and exclusion clauses. Even when the law imposes mandatory rules that cannot be excluded by agreement (such as in consumer protection laws in Nordic countries and especially product liability regulation deriving from the EC Product Liability Directive (85/374/EEC), which command instead of serve the contracting practice), there still is room left to alter the legal status of the social behaviour in a way that circumvent the indispositive rules. Contracting parties can structure their relationship in ways that avoid the application of these rules[924].

In the software business such an effort is made by stating that software is licensed for use, i.e., a limited right to use is given, not sold and that no sale of goods occurs and no contracts for sale is drafted. If this argument is accepted in courts, which is largely an unresolved issue in many jurisdiction, the default liability rules both in general and consumer contract laws and in specific product liability

---

from its earlier precedents. In practice, however, precedents of the Supreme Court are followed in cases arising after the precedent has been given and involving a similar point of law.

[924] Contracting parties can act according to a contract that would be considered invalid under the law, apply contractual terms that are against indispositive rules, or even breach contractual obligations without consequences, unless neither of the parties oppose to this or have a possibility or capability to do this. As a regulatory mechanism, therefore, contracts have the potential to regulate contract law itself in the sense of controlling and modifying its application; even in relation to the mandatory (indispositive) consumer contract rules as Hugh Collins argues in Regulating Contract Law, p. 16-17. This is an extension to the above mentioned traditional perspective in regulating contracts. In the Nordic doctrine the restrictions in the guidance functions of contract law have been recognised, e.g., by Jan Hellner in *Lagstiftning inom förmögenhetsrätten*, p. 162-164. Note that this is often considered to be infeasible as a conscious contractual technique at least in standard forms because it exposes the vendors to the reactions of the consumer authorities and other controlling organisations, and a plainly illegal contractual practice might lead to severe damages to reputation if it becomes public. In general terms, this has been pointed out, e.g., by Mika Hemmo, *Sopimusoikeus III*, p. 16-17, in considering the limits to the freedom of contract in his textbook on the technique and tactics of contracting.

rules following the EC Product Liability Directive (85/374/EEC) are inapplicable since they mainly apply to the sale of goods and require the transfer of ownership[925]. Instead of transferring ownership to the users, which is a prerequisite of applying the sale of goods rules, the licenses only give a right to use the product (software is licensed for use, not sold).

Despite the equivocal nature of this rather established contractual practice in the software business[926], it seriously affects the incentives for secure software development and informing customers about the vulnerability of their software. It not only makes the default dispositive rules to look as inapplicable, but also the mandatory rules that otherwise cannot be contracted to the disfavour of the buyer or the injured person prior to the damage.

> Following the Finnish doctrine we can now separate three different situations in the applicability of liability disclaimers. 1) If the contract made about the transaction of COTS software is binding on the parties, then so are the disclaimers[927]. Whether or not the transaction is a license for use or a contract for sale has no relevance in this situation. 2) If the software contract is not binding, the transaction is a contract for sale instead of a license for use. The default rules

---

[925] The validity of the licenses, especially in their mass-transaction form, is still a matter of dispute and no clear priority has been given to the terms of the license in deciding whether the provision of the software is that of license or sale. Instead, the consideration typically goes by the factual characteristics of the situation in question. Especially if no time limits for the license has been provided and it is provided upon single payment instead of continuous license fees, then the contract is more easily considered as the sale of goods than as the license of the right to use. However, it has to be stressed that this is far from clear as noted by Mika Hemmo, Vuoden 2000 ongelma ja siviilioikeus [The year 2000 problem and civil law], p. 25-27, in relation to the Finnish doctrine.

[926] The effectiveness of this strategy under various jurisdictions is uncertain especially in relation to consumer contracts in COTS software and, to an increasing degree, also in commercial contracts where a contracting party is at a weaker position.

[927] Naturally there might be the possibility for courts to invalidate the terms in individual disputes or to forbid their use in the future.

> of defectiveness and other product liability rules apply. 3) If the liability disclaimers and exclusion clauses in software licenses are found by the court to be binding, but the transaction is found to be a contract for sale, then the liability depends on the position of the buyer. If the software is bought for business purposes the disclaimers are binding. If the software is bought for private use by a consumer, the interpretation of the transaction to be a contract for sale makes the indispositive consumer protection rules applicable. In this case it is possible that the liability disclaimers are not considered to be binding.

This further strengthens the already strong bargaining power in post-breach negotiations created by standard form contracts. When a complaint is made, the vendor can point to clauses that exculpate or greatly diminish the responsibility of the business, which serve to discourage the pursuit of the complaint[928]. This further reduces the buyer's willingness to litigate over to contractual terms and dispute about the defectiveness or insecurity of the software even though grounds for argument could exists. The legal shields the buyer has to break are strengthened in practice even though their validity might be questionable at least under several consumer protection laws.

### 5.3.3 Implementation

Like most private law rules, the product liability rules are typically implemented via ordinary civil proceedings. Normal sales law remedies connected with the defective goods (conformity with contract) provisions discussed above are available both to individual and industrial buyers and the injured party has a possibility to seek compensation for damages under specific product liability rules and general tort law.

This is characteristically a form of self-implementation. Application of the rules is not watched over by any prosecuting or supervising

---

[928] Collins, *Regulating Contracts*, p. 229.

authority[929]. Initiative for the process of applying the rules is left to the parties themselves, or more specifically, to the party whose interests have been violated[930]. The party wronged by a breach of contract or by an extra-contractual tort can decide whether to address any claims to the other party or the causer of the damage on this score and whether she wishes to employ legal mechanisms in her attempt to have these claims met. She has the discretion whether or not to enforce the rules through an ordinary court procedure or to negotiate a revision of the standards where possible.

This discretionary power exposes the implementation via court proceedings for several hindrances. The injured party has to have self-interest to take the issue to the court. In relation to software security issues there are several shortcomings to this self-interest not just for the individual consumers, but also for the industrial buyers. Some of them are general to the type of self-implementation via ordinary civil proceedings typical to private law and others specific to the software industry.

One assumption behind the effectiveness of this self-implementation is that consumers and business purchasers, who are allocated the task of inspecting their purchases for quality defects and of identifying the damaging software, can actually do so. Even though this applies to most search and large part of experience characteristics of goods, in relation to credence characteristics (i.e., latent defects which might only be detected as a result of inspection by persons with technical expertise such as software vulnerabilities) it becomes extremely difficult. Even users of the software simply lack the

---

[929] There are exceptions in the case of consumer protection, such as in the case consumer marketing regulations in Finland. Stig Strömholm makes the same notion in relation to the implementation of indispositive private law rules in Sweden in *Rätt, rättskällor och rättstillämpning*, p. 158. These will be discussed below.

[930] As noted by Hugh Collins, *Regulating Contracts*, p. 64, in relation to the regulation of contracts, monitoring compliance with the quality standards set in private law and contractual clauses is delegated to the parties themselves.

sophistication to police detailed technical features and even access to such expertise is limited due to the prohibitive costs[931].

Lack of knowledge in software vulnerabilities, even access to essential information and expertise has been emphasised many times above. This is a real problem not just for the customers but also for the judges. This results in difficulties in proving defectiveness and causation even in strict (no-fault) product liability cases in consumer goods[932]. Problems in detecting the lack of quality and security in

---

[931] Collins, *Regulating Contracts*, p. 298, correctly suggests in relation to the contractual regulation of quality that in the case of credence goods, public regulation, where a substantial administrative apparatus is required to examine samples from vendors, seems to be the only efficacious method to monitor conformity. Certainly, selective regulation of potential design faults that requires some components to conform to detailed technical specifications monitored by a regulatory agency could play a certain role in the regulation of secure software development. This is so despite the expenses related to such a model resulting, e.g., from the need to use a team of inspectors examining samples from vendors for their software products in order to determine whether they meet the quality standard. However, the development of such regulatory standards that would apply to different types of software used in a variety of different context with altering security requirements has proven to be problematic. The speed of change in the software business does not ease the development of regulatory standards that would be stable and detailed enough for the regulatory agencies to monitor. As a result, such a model of public regulation might be applicable only in safety critical contexts and in relation to critical infrastructures where human life is at stake.

[932] As Reimann, Product Liability in a Global Context, p. 150, notes in analysing product liability rules worldwide in 2003, strict manufacturer liability is often not much of an advantage for the injured parties. The plaintiff must still show defectiveness and cause, and not having to prove fault is frequently a small benefit at best. Liability has often been strict, or near-strict, under generic contract or tort rules in practice anyway. It must be emphasised that traditional tort or contract liability is not absolute. But neither is the liability under the EC Product Liability Directive (85/374/EEC). The consumer expectation test as well as the defences recognized in the Directive creates considerable loopholes. It is evident that questions relating to the burden of proof continue to be controversial, and are seen by many to be of real practical significance. According to the perception presented by some of the consumer representatives (i.e., consumer associations both at national and EU levels) in a survey study on product liability in the EU conducted by John Meltzer, Rod Freeman and Siobhan Thomson, *Product Liability in the European Union*, p. 47,

software combined with the consequent problems of proof about the insecurity and defectiveness of product in the case of security is a serious hindrance for implementation. Difficulties in proving that a violation of a regulatory standard has occurred negatively influence the decisions of injured parties to take the issue to the court.

In general contract and tort law typical starting point is that those who allege fault have to produce the initial evidence. Strict product liability rules resolve this problem in the detecting and proving that a breach of standard has occurred, by transferring the burden of proving compliance onto the alleged violator. When the courts regard proof of an unsatisfactory product quality and unreasonable product security as sufficient evidence of breach of the regulatory standard, the legal process can imitate public regulation by in effect requiring the vendor to demonstrate the absence of fault. As Hugh Collins notes, the capacity of private law to detect failures of regulatory compliance turns crucially on such techniques for the reversal of the burden of proof[933].

However, as Mathias Reimann notes in analysing product liability rules worldwide, even strict manufacturer liability is often not much of an advantage for the injured parties. The plaintiff must still show defectiveness and cause, and not having to prove fault is frequently a small benefit at best.[934]

---

consumers are unfairly disadvantaged by the burden of having to prove defect and/or causation in product liability claims. The concern mainly arises from perceived difficulties in proving claims due to a lack of legal or other resources needed to investigate them properly, or to an inability to gain access to essential information. Such problems are seen to be particularly acute in relation to technical products, or where the alleged injuries are of a complicated nature.

[933] Collins, *Regulating Contracts*, p. 89-90.

[934] Reimann, Product Liability in a Global Context, p. 150. Liability has often been strict, or near-strict, under generic contract or tort rules in practice anyway. It must be emphasised that traditional tort or contract liability is not absolute. But neither is the liability under the EC Product Liability Directive (85/374/EEC). The consumer expectation test as well as the defences recognized in the Directive creates considerable loopholes.

It is evident that questions relating to the burden of proof continue to be controversial, and are seen by many to be of real practical significance. According to the perception presented by some of the consumer representatives (i.e., consumer associations both at national and EU levels) in a survey study on product liability in the EU conducted by John Meltzer, Rod Freeman and Siobhan Thomson consumers are unfairly disadvantaged by the burden of having to prove defect and/or causation in product liability claims[935]. The concern mainly arises from perceived difficulties in proving claims due to a lack of legal or other resources needed to investigate them properly, or to an inability to gain access to essential information. Such problems are seen to be particularly acute in relation to technical products, or where the alleged injuries are of a complicated nature.[936]

In addition to the lack of information about the security and quality of software together with the shortness in the access to expertise, also another information problem certainly inhibits litigation. The still relatively excessive uncertainty as to the very basis upon which a court may decide certainly furthermore inhibits litigation. This emphasizes in international transactions since the choice of law rules are by no means clear even in relation to consumers. When not only the success of the claim is uncertain but already issue of jurisdiction and choice of law together with the applicable law and its basic interpretations too, there is bound to be unwillingness to invest high litigation cost on such uncertain ground[937]. Even though there are written laws with

---

[935] Meltzer et al., *Product Liability in the European Union*, p. 47.

[936] Despite the acknowledgement of these problems in the burden of having to proof defect and/or causation in product liability claims, and the considerations of the ways to mitigate the burden of the injured party, John Meltzer, Rod Freeman and Siobhan Thomson in their report to the European Commission on product liability in the European Union still do not recommend changes to the directive. Of the different alternatives to mitigate the burden of proof of the injured party, see also Mononen, *Yritysten välinen tuotevastuu*, p. 28-29.

[937] Meltzer et al. in *Product Liability in the European Union*, p. 34-37, provide evidence of product liability claims in the EU becoming more successful in the past 10 years. The factor that was most commonly identified by the participants to their questionnaire as contributing to the increase in success of product liability claims was greater access to legal assistance/advice. A

accompanied drafting material and guiding case law on different types of goods that provide relative certainty in the area of consumer product liability rules, in the case business relationships written laws are scarce and, when exist, typically dispositive and open to interpretations.

When even some of the most basic questions concerning the application of provisions of contractual and non-contractual liability in the software business admit no easy or certain answer in many jurisdictions, this uncertainty becomes excessive. Especially the cases concerning COTS software are problematic mainly due to the above mentioned use of liability disclaimers in licenses, the power of liability disclaimers in post-breach negotiation, several layers of intermediaries that may break the causal link between a defect and a damage[938], and the difficultness of specifying cause to a certain software product in a system consisting of several highly interacting components from several vendors and of which some could even be developed in-house. The possibility of vendors of defective software to raise a strong argument that the plaintiff-purchaser who has failed to apply patches or to take other simple self-protective measures (such as using firewalls or virus-scanning software) to the sufficient degree has been contributory negligent by no means lower the threshold for litigation[939].

---

significant number of participants also identified the Directive as contributing to the success of product liability claims. Harmonised rules provide a clear basis for litigation.

[938] As Jennifer Chandler explains in *Improving Software Security*, p. 10-11, security-related software flaws on their own typically cause no losses. There can be possible losses due to the need to repair the defective software even though the vulnerability in it has not been abused. However, the losses are primarily inflicted when a third party deliberately takes advantage of the vulnerabilities. Of course the action of retailers, implementers such as IT consultants helping in the strategic procurement and implementation, system integrators and support firms providing training or technical support, are also important.

[939] As Jennifer Chandler, *Improving Software Security*, p. 12, points out in relation to the U.S. tort law, the plaintiff-purchaser cannot be held contributory negligent for failing to apply patches to their software in situations where the attack is launched before patches are ready. Still, the failure to take other self-

Another issue to consider is the costs of implementation. As Hugh Collins notes, the absence of public financial resources to support regulatory compliance tends to limit the ability to enforce the rules to those with considerable wealth[940]. When the costs of litigation in an ordinary civil court are often high[941], the enforcement of those security standards where the damage caused by the violation imposes only minor costs on the plaintiff, i.e., the user of the software in the case of insecure software, is often deterred. Users of the software are reluctant to incur the expense of litigation in order to obtain the small benefit of compensation or return of the price.[942]

It is important to note that these problems do not necessarily lead to suboptimal enforcement from the regulatory perspective. For example, the lack of self-interest among private individuals (consumers) due to the slight financial losses compared to the uncertainty and costs of litigation is not present to the same degree

---

protective measures may still constitute contributory negligence. However, the degree to which average Internet users can be expected to be able to patch their systems cannot be high. There simply are too many patches from too many sources for the variety of software in use in a typical personal computer. It would be unreasonable to expect home users to have the time and to be capable of installing all the patches, at least as long as automated patching is not sufficiently available and reliable enough.

[940] Collins, *Regulating Contracts*, p. 87.

[941] As Reimann notes in Product Liability in a Global Context , p. 152: "In most countries, a victim's access to court is also comparatively expensive [to the U.S. system: author's note] because their law does not permit a (full-fledged) contingency fee system; without such a system, a plaintiff faces potentially high up-front retainers as well as the burden of other litigation costs. In addition, suing is often financially risky because in the majority of systems, the loser has to pay not only his own attorney's fees and but also the other side's lawyer; this doubles the cost risk and militates against going to court".

[942] Note that the costs of the private law type of monitoring compliance with regulation is considered to be much cheaper in total than the public administrative control model of public regulation. However, the relevant issue in here is that the costs to the individual who ponders upon the filing of a complaint are high. To her, the overall cheapness of this type of monitoring is not relevant. She only sees the private costs of litigation.

in the case of industrial buyers. They have a clear self-interest in taking defective product cases into court. The vulnerabilities found in the components of their systems not only can cause substantial damages when the risk materializes, but can also hamper the sales of their own system using the software as a component or damage the public image when the existence of the vulnerability also in their systems (that uses the vulnerable component) becomes widely known.

However, in COTS software business the litigation threshold seems to be relatively high. In addition to these cost increasing and litigation threshold heightening factors there are hindrances for litigation that stem from the characteristics of the software business. Due to the rapid evolution of the business, companies do not have time to look behind or ask for financial compensation in cases where the component has not been delivered, lacks in quality, or produces risks. Software companies are not eager to start fighting over broken or unmet contracts[943]. Keeping pace with the development is considered as such enough demanding and resource consuming. Also the reputation damaging effect of excessive litigation might influence the desire of networked businesses to start litigating. Litigation becomes

---

[943] The speedy COTS software business seems to lack interest to take the issue to court. This approach in COTS software business is verified in an empirical research by Juhani Warsta in *Contracting in Software Business*, p. 139. This lack of interest is present in COTS software business despite the weak connection between the vendor and the buyer. The spot focus of the relationship could lower the threshold for litigation in theory, since there hardly is a comparable fear of losing a strategic partner as in more relational MOTS or tailored software businesses. Note that longer term relationships and relational contracts are used with channel operations with the resellers and in systems integrator partnerships (Warsta, *Contracting in Software Business*, p. 181). Even when separate contracting outside standard forms occurs, the incentive to emphasise contractual entitlements is low due to its reputation damaging effects as Hugh Collins emphasizes in *Regulating Contracts*, p. 174, with extensive references in relation to contracting in general. The desire to maintain a cordial ongoing relationship is not present to the same degree. In longer term relationships the quest for legal remedies is not expected to be high due to the primacy of relational and economic normative frameworks of contractual behaviour compared to the contractual in the terms used by Collins in *Regulating Contracts*, p. 128.

an option only when the long-term relation does not seem to be practicable anymore and the parties feel unable to achieve an accommodation that preserves some diminished benefits for both[944].

The costs structure for the consumers also changes when certain jurisdiction, such as Finland and Sweden, provide with the possibility of taking product liability claims into consumer complaints boards. This offers consumers an accessible and inexpensive forum that lowers the threshold of litigating. Even though the decisions are not binding, they are usually followed by the parties to whom they are directed at the specific jurisdiction in question. However, when the national implementation of the EC Product Liability Directive (see art. 9 (b)) does not allow a consumer to sue for the first 500 Euros of damage, this minimum threshold has a real impact on consumers[945]. The damage suffered by an individual consumer due to insecure software can often remain at a lower threshold. The force of such non-binding national rulings can also be minimal when the business is essentially international and the issues to be solved highly speculative[946].

---

[944] Collins, *Regulating Contracts*, p. 329-339.

[945] The availability of an inexpensive and readily accessible tribunal would otherwise make it economically viable to bring relatively low-value product liability claims to the Consumer Complaint Board. The negative impact of the lower threshold (€500) in the article 9(b) of the EC Product Liability Directive in Finland is pointed out also in a recent EU study on product liability by John Meltzer, Rod Freeman and Siobhan Thomson, *Product Liability in the European Union*, p. 50.

[946] EU regulators have responded to the problems in the enforcement of consumer protection laws at the internal market level. Directive 98/27/EC of the European Parliament and of the Council of 19 May 1998 on Injunctions for the Protection of Consumers' Interests (the Injunctions Directive), OJ L 166, p. 51-55, establishes a common procedure to allow a qualified body from one Member State to seek an injunction in another. Regulation (EC) No 2006/2004 of the European Parliament and of the Council of 27 October 2004 on Cooperation Between National Authorities Responsible for the Enforcement of Consumer Protection Laws (the Regulation on Consumer Protection Cooperation) OJ L 364, 9.12.2004, p. 1-11, which applies from 29 December 2005, removes barriers for information exchange and cooperation and empowers enforcement authorities to seek and obtain action from their counterparts in other Member States. It is noteworthy that the Regulation does

The enforcement could become more effective when the possible legal avenue of seller liability under contract is used[947]. When the seller has the possibility to make backward demands towards the channel operators and the software house that developed and shipped the insecure product, this could provide a more capable and both financially and informationally better equipped party to demand more secure software and to take the issue to the court when necessary[948]. Also the national legal solutions might be more effective against the retailer operating in national markets; at least there is a suitable defendant to sue without invoking complex and vague choice of law rules[949]. However, the EC product liability rules for consumers are built for direct litigation possibilities[950]. Mathias Reimann identifies

---

not seem to apply to the enforcement of the Product Liability Directive (85/374/EEC).

[947] Even though the doctrine of privity of contract (or the principle of relative effect) usually confines the plaintiff to a suit against the immediate retailer, this doctrine has been modified in some jurisdictions in order to enable a direct action between the consumer and the manufacturer for defective quality and security also in product liability cases (Meltzer et al., *Product Liability in the European Union*, p. 16-17). The direct contractual relation between the user and the vendor without retailers in the between is often achieved through private ordering in the form of manufacturer's guarantees, which in effect promise the cost of repairs directly to the consumer (Collins, *Regulating Contracts*, p. 301). In the COTS software business the same result is achieved by the use of EULA's.

[948] As Hugh Collins notes in *Regulating Contracts*, p. 301, the retailer may be in a more powerful position than the consumer to threaten the use of non-legal sanctions, such as an unwillingness to stock a particular product in future, so that the liability rule has a powerful supplement.

[949] Robert Bradgate in a special feature of the Journal of Information, Law and Technology concerning the Y2K problem, Beyond the Millennium, heading 3.3.4 last sentence, argues on practical and policy reasons in the English legal system for the issuance of liability for defective COTS software in a consumer context primarily on the retailer. At the same time he argues against the interpretation of the software user license (EULA) as establishing contractual privity between software manufacturer and user.

[950] Under EC Product Liability Directive (85/374/EEC) the primary liable parties are the manufacturer or parties that give the impression to be the manufacturer. These include, e.g., party importing the product to the EC markets or a party presenting herself as the producers by putting her name,

the impossibility to sue anyone but the producer (and his alter egos) as one of the serious disadvantages of the EC Product Liability Directive (85/374/EEC) and the remedies fashioned after it also in many non-European countries compared to general tort and contract claims[951].

Due to the concurrence of both contractual and non-contractual (delictual) bases for liability in the case of vulnerable software the plaintiff can avoid many of the caveats of specific product liability rules by using the right to choose the legal category which will give the most advantageous outcome. However, this would require extensive knowledge of the disadvantages and benefits of different legal avenues, which is not generally available. Understanding of the wide legal basis for product liability is still relatively scarce even in general, not to even speak about specific issues of the software business.

In light of these difficulties faced by the software user in bringing out a claim against the vendor, Jennifer Chandler makes an interesting suggestion for these implementation level problems. Instead of depending upon a purchaser of defective software to be the plaintiff in a hypothetical lawsuit against the developer of unreasonably insecure software, Chandler argues that a more promising suit could be brought by victims of DDOS attacks. As plaintiffs they are not

---

trade mark or other distinguishing feature on the product (article 3 of the EC Product Liability Directive). This emphasis the role of contractual product liability rules also in consumer relations, since it is the seller of the software, to which the consumer has the most direct connection with when the software is delivered via the retail market and not purchased directly from the manufacturer via the Internet. At the same time there is a need to consider security issues also under the indispositive default rules for defective products in general contract law or doctrine. If a software product is not secure enough, then there ought to be a possibility to invoke contractual rules for defective products. Due to the freedom of the plaintiff to choose the legal basis under which to bring a product liability case into a court, this already is possible in many European jurisdictions. This possibility is, however, of value only in those situations where the software product has been bought from a retailer and not directly from the developer.

[951] Reimann, Product Liability in a Global Context, p. 150.

open to charges of contributory negligence as there is essentially nothing they can do to protect themselves. They suffer the kind of concentrated loss that would make litigation attractive. In addition, they do not face the obstacle of contractual disclaimers and limitations of liability within the software licence agreements that exist between software vendors and the owners of the insecure computers used to launch a DDOS attack.[952]

Even though the target of a DDoS attack might be the best suited plaintiff in a lawsuit against the vendor of insecure software, there is no reason to withhold from improving the institutional support for the suits brought by the users of insecure software. At least they have a right to compensation, even if the directive force of liability rules would be sufficient in the current situation where only the victims of DDoS attacks can effectively claim for damages. In addition, software vendors are by no means the only parties that could be held liable for DDoS attacks. A more likely alternative target for liability suits for the victims of DDoS attacks could be found from those who maintain the vulnerable machines and a party with even deeper pockets could be found from the service providers that relay the attacks trough the networks.

Even though the enforcement of private law regulation seems to lack resources, this does not necessarily lead to a failure in achieving an optimal level of enforcement. Compliance with the regulatory standards can, and often does come also from other places than from the enforcement of the rules. As Collins notes, the enforcement mechanism provided by the law is usually a minor consideration in the incentives for regulatory compliance[953]. As noted above, regulatory compliance does not primarily rely on legal sanctions, but rather upon

---

[952] Chandler, Security in Cyberspace, p. 234. Chandler argues in Security in Cyberspace, p. 258, that resulting greater attention to security at the software development stage would have the positive side effect of reducing insecurity generally even though lawsuits are most likely to be successful only in the context of DDOS attacks. This is because the same types of vulnerabilities that lead to DDoS attacks also lead to other nuisances such as the circulation of malicious code in general.

[953] Collins, *Regulating Contracts*, p. 88.

kinds of non-legal sanctions such as refusal to do business and damage to reputation[954] and associated costs that are not directly linked to the legal liability.

The refusal to do business is a powerful sanction where the deceiver derives much of its business or income from the other contractor. It is most powerful in long-term relations between enterprises. In relational contracts the key incentive to perform contracts faithfully, particularly with regard to quality, derives from the greater benefits to be achieved from the income from a stream of future sales rather than one-off benefits from defaults[955]. In the case of COTS software being used as a component in an information system, the refusal to do business can become a powerful sanction especially when the buyer is sufficiently large and influential such as the government[956].

Note that the communicative function of the adjudication process is at its strongest. When the constitution of a market is considered to depend heavily upon business reputation as a source of trust, then the authoritative declaration by a judicial organ or any institution that possesses the respect of the trading community (e.g., arbitrator, consumer ombudsman) serves the function of collecting and disseminating what participants in the market regard as reliable information about trustworthiness. Participants in the market are then

---

[954] This emphasizes where the information about default can be disseminated rapidly to other participants in the market as Collins notes in *Regulating Contracts*, p. 114. This can be the case in relation to COTS software. Note that vendors can, however, use license terms restricting product comparisons to target such dissemination of information.

[955] Collins, *Regulating Contracts*, p. 114.

[956] Government procurement has for this reason been seen as one of the regulatory tools that could be used to influence secure software development. For example, the NCSP (National Cyber Security Partnership) in *Technical Standards and Common Criteria*, Appendix B, p. 2, recommends that the U.S. Government should require vulnerability analysis of products as a separate condition of procurement or part of the Common Criteria. The capacity of government to influence secure software development via its procurement policies calls for a separate analysis in subsequent studies.

better able to adopt the typical non-legal sanctions to the merchant whose reputation has been so damaged.[957]

Despite the force of non-legal sanctions there still is a pressing need to address the problem of the lack of resources inherent in the design of the private law self-enforcement mechanism in order to improve the current level of security in software[958]. The actual level of compliance with standards in the absence of legal enforcement appears to fall below an optimum level as has been argued above[959]. The non-legal sanctions and the costs associated with insecure software have not been sufficient to improve the level of security in software[960]. Also the enforcement mechanism appears to provide weak additional incentives, especially when the possibly available cheaper enforcement alternatives like ombudsmen and consumer complaint

---

[957] Collins in *Regulating Contracts*, p. 124, considers this communicative function to be more important than the legal sanctions attached to the breach of contract.

[958] According to Hugh Collins, *Regulating Contracts*, p. 88-9, the situation where there is a need to address the problem of the lack of resources inherent in the design of the private law self-enforcement mechanism is present when compliance appears to fall below an optimum level, and the enforcement mechanism appears to provide weak additional incentives. I.e., the question of the optimum level of resources to be spent on the enforcement must be assessed in the light of information regarding actual levels of compliance with standards in the absence of legal enforcement.

[959] This is the argument made especially in chapter 2.

[960] The costs associated with non-legal sanctions have not been foreseeable enough to provide sufficient incentives for secure software development from the societal point of view. One of the main non-legal sanctions, i.e., damage to business reputation is not all that feasible in relation to credence characteristics like the vulnerability of software. At the same time patching provides a reasonably low costs alternative to product recall and repair as has been argued above. Customers have also tolerated the vulnerabilities due to several historical reasons and for their lack of capability and information to compare products in terms of security. Consequently, shunning has not been a real risk for software vendors, even though the climate is changing. Customer tolerance of insecurity might, however, be vanishing over time. It remains to be seen whether this is sufficient to improve the security of software in a sufficient manner but, as the theoretical considerations presented above predict, it is not likely to do so to the level sufficient to the networked society as a whole.

boards are in a relatively weak position to influence international software houses[961].

Another type of implementation of product liability rules is via persuasion, cooperation and negotiated solutions which are used instead of punishment of deviant behaviour. Contractual negotiations between the customer and the COTS software vendor together with the negotiations between the representatives of the consumer interests and the vendors are central implementation measures. This holds the promise for better internalisation of the standards in the enterprises' decision-making concerning secure software development and wider acceptance of the rules because the juxtaposition and conflicts between the parties are not emphasised. The implementation does not originate from the assumption that the economic burden related to the threat of sanctions and the related moral blame is necessary in order to get the objects of regulation to comply[962].

Consumer redress for poor quality and security COTS software, like any other good, operates in practice almost entirely outside the legal system. It is handled usually by complaints procedures of manufacturers and retailers, which differ in their likelihood of producing redress for the consumer. Traditionally, the software industry has not succeeded well in this[963]. There are several feasible explanations.

Already the above mentioned weaker informational position of the customers (both private individuals and industrial customers

---

[961] Product liability as a form of consumer protection is problematic from the point of view of enforcement in the network level. The insufficiency of national laws and their enforcement and the need for network level implementation is emphasised in the characteristically internationally COTS software business.

[962] This juxtaposition of the parties is a weakness of the implementation strategy basing on the threat of sanctions and their enforcement (deterrence), as raised in sociolegal analysis of regulation such as Ayres and Braithwaite, *Responsive Regulation*, p. 25.

[963] The statistics gathered by Cem Kaner and David Pels in Software Customer Dissatisfaction almost a decade ago already testify this.

including competitors) with their possible representatives like consumer organisations in relation to security related vulnerabilities, causes problems to this cooperative model of self-implemen-tation[964]. Further problems for the cooperative implementation are caused by the same above mentioned reasons that lie behind the ineffectiveness of the deterrence model of enforcement.

The success of cooperative enforcement depends on the possibility to detect and correct or compensate the situations where the level of security is not high enough. When there are severe hindrances for litigation and the possibilities to find out the low level of security and quality or to use that knowledge, even when available, to negotiate higher security standard next time, the COTS vendors might not be sufficiently concerned with future encounters with the same customers to resist the short-term temptation to minimize the development costs at the cost of security. Harm causer will be likely to be able to settle out of court for negotiated sums that are lower than those that would create efficient levels of deterrence also because of the strong post-breach bargaining position resulting especially from the seemingly effective use liability disclaimers and exclusions clauses by the software vendors[965]. The negotiation force of the customers alone might not sufficient to shift the current balance under which COTS software

---

[964] Even though industrial buyers are in a better informational position to evaluate the products and vendors claims, they still are at the mercy of the COTS software vendor largely because the lack the means (not only are white-box methods largely unavailable, but also the tools for testing even on the developer side are relatively undeveloped and not in wide use), financial resources, and sufficient incentives to evaluate the products security.

[965] The settlements, especially when done confidentially so that everything from initial discovery through the actual details of a settlement are kept secret, can lead to lower average quality inputs used and lower average security of products sold, than that which would be produced if a firm were committed to openness. This has been shown by two U.S. law and economic scholars Andrew F. Daughety and Jennifer F. Reinganum in Secrecy and Safety, p. 1074-1091. The final applicability of this argument and other conclusions made about the effects of the choice between confidential or open settlements to the settlements in software product liability cases is, however, uncertain and requires context specific research in the future.

vendors can get away with several security related vulnerabilities in their systems.

*Specific implementation issues in the duty-to-disclose rules.* Since the legal duties to disclose are somewhat based on different rules that are only partly complementary to the product liability rules, a few separate notions on their implementation is in order. The variety of material legal rules establishing duties to disclose vulnerability information, varying from rules on general unfair marketing practices (such as unfair advertising) and unfair competition to the "lack of safety" standards in product liability law and general contract and tort law can naturally be implemented in various ways. Even in culturally and economically relatively homogenous area like the EU the variations are extensive; some rely on competitors as primary interveners of unacceptable practices, others on consumer or similar organisations, self-regulatory bodies, individuals as parties in court proceedings concerning injunctions or damages, and even on administrative control[966]. Due to the large variety in implementation methods the analysis or their effects cannot be detailed or in depth. Only cursory notions on general level can be made[967].

In the case where the disclosure rules are based on legislation on marketing practices, that in many countries (like the Nordic) concern unfairness both from the point of view of consumers and

---

[966] The need for the participation of several actors is considered obvious as noted by Thomas Wilhelmsson, Administrative Procedures for the Control of Marketing Practices, p. 142-143 and 146, in relation to the regulation of marketing. The different approaches to the regulation of marketing inside EU has been highlighted also in the European studies on better regulation, such as the Better Regulation Task Force's (renamed as Better Regulation Commission in the beginning of 2006) study on the use of alternatives to classic regulation in the EU, *Routes to Better Regulation*, Annex 2.

[967] It is important to note that the implementation of these rules is partly also based on similar self-implementation via ordinary civil proceedings as product liability rules. Thus, also the above considerations apply to their implementation. Below, only those types of implementation that differ from the product liability rules are considered.

competitors[968] the function of a watchdog is entrusted to competitors. Competitors can seek prohibition for unfair marketing from courts (in Finland from special Market Court that also applies the similar rules in the Consumer Protection Act). Even though competitors on the average are more aware of their rights and can be expected to be more knowledgeable of the security issues, they might not be incited to take unfair marketing practices to the court. This is so especially when they, as is the case of security related vulnerabilities in software, are subjects to the same problems and might even conduct similar marketing practices. In nearly monopolistic situations present especially in operating systems software, there might not even be competitors that could intervene against the marketing.

Consumers incentives and possibilities to litigate under unfair marketing practice provisions, where provided a possibility to do so, is even more constrained, e.g., due to the lack of information, fear of costs, small losses to the individual compared to the consumers as a collective, and evidentiary uncertainty. Even when collective level intervention, that is apart from the individual disputes, is enabled, as is done in many EC countries (e.g., consumer organisation or similar bodies are empowered to institute injunction proceedings against unfair marketing practices) the results might not be sufficient when a complex technological issue like the disclosure of vulnerabilities is at stake. The organisations lack the economic resources, expertise, and incentive due to the weak mutual understanding between consumers and possible conflicting interests (others, such as early adopters may favour new innovative features more than security compared to average users that have no use for such features).[969]

---

[968] In Finland, where a special legislation on unfair competition (Finnish Unfair Business Practices Act 1061/1978) concerns practices that are unfair to other entrepreneurs, even it is presented as having an indirect protective function also in relation consumers. This is explicitly stated in the Government Proposal 114/1978 and is repeated in the literature such as Castrén, EU-Suomen markkinaoikeus, p. 237. As noted above, Finland has also regulation on unfair contract terms that covers both enterprise and consumer relations.

[969] Same argument in general is made by Thomas Wilhelmsson in Administrative Procedures for the Control of Marketing Practices, p. 147-148.

Another form of implementation that is used in relation to the vague general marketing clauses forbidding unfair marketing and unfair contract terms is the use of administrative control system that in the Scandinavian countries is employed by the consumer ombudsmen[970]. The public authorities have a central role in supervising marketing to consumers and can conduct negotiations with the enterprises and their organisations. In the Nordic countries, the ombudsman also acts as a plaintiff in injunction proceedings against entrepreneurs in the special courts.[971]

This form of implementation through administrative control, where the public authority bases its work on a set of broad general clauses on unfair marketing issued in legislation and the possibility to issue injunctions against unacceptable marketing, has certain benefits that could make it influential in relation to the disclosure of security related vulnerabilities in software. The vague general legislative mandates enable the public authorities to tackle with multitude of issue of which the disclosure of vulnerabilities in software could be one even though it currently is not in the agenda[972]. By mainly using negotiations with the enterprises that behave in an unacceptable way (in this case, do not inform sufficiently of the security of their products) and relevant business organisations in order to create concrete and acceptable standards for marketing behaviour, the internalisation of the standards provided into the decision-making of the enterprises is made more

---

[970] Similar supervising administrative agencies exist, for example, also in UK in the form of the Director General of Fair Trading and the Office of Fair Trading. See the web pages of the Office of Fair Trading at http://www.oft.gov.uk/ [20.1.2006].

[971] Wilhelmsson, Administrative Procedures for the Control of Marketing Practices, p. 143-144.

[972] There is not a long way from the actions taken by the Finnish Consumer Agency in advising the providers of broadband connections to inform their customers of the need for security measures like firewalls, anti-virus software and updated operating systems in the open networks, to advices given to software vendors to inform about the vulnerabilities in their products. Of the previous recommendation see the Press Release 24.11.2004 of the Finnish Consumer Agency at http://www.kuluttajavirasto.fi [23.2.2006].

likely. This also helps in striving at a broad acceptance of the principles, which is a prerequisite for efficiency that reaches further than the single control measure and decisions.[973]

However, since the emerging COTS software business is not particularly well-organised as a community, despite the powerful lobby groups like the Business Software Alliance representing the big commercial enterprises, and is essentially international in nature, the national administrative control is not likely to be interested in software vulnerability issues anytime soon[974]. Even if the security related software vulnerabilities are raised as an important consumer issue, the national administrative control system runs into problems in countries whose software is mainly imported or bought directly from abroad, e.g., via the Internet[975]. Due to the complexity and newness of the issue the administrative control system is not likely to react anytime soon when even the major players in the field are not aware of the vulnerability issues.

In sum, due to the self-implementing nature and many shortcomings of the private law implementation of the legal duties to disclose and of the administrative control model, the deterrence model of enforcement seems ineffective in relation to secure software development. The expected costs of non-compliance are not increased

---

[973] Wilhelmsson, Wilhelmsson, Administrative Procedures for the Control of Marketing Practices, p. 155.

[974] Interests in information security concern much more traditional consumer protection issues like the protection of personal data and payments in e-commerce. For example, the Finnish Consumer Ombudsman in instructs for Internet marketers states clearly how the transfer of personal data and the electronic payments are secured. The common position of the Nordic Ombudsmen is that if the transfers or payments are not sufficiently protected, then the marketer has to inform the consumer about this. See the web pages of the Finnish Consumer Authorities (agency and ombudsman) at http://www.kuluttavirasto.fi [23.2.2006], also in English.

[975] Both the administrative control system and the self-implementation of the legislation concerning marketing practices lose much of their efficacy when a consumer deals directly with a foreign vendor. As pointed out by John Rothchild in Co-Regulating the Internet, p. 180, attempts to enforce the laws of the consumer's jurisdiction face both legal and practical roadblocks.

to a degree that would turn the calculus assumption behind the standard deterrence model in favour of compliance due to the several shortcomings of the implementation. The value of expected costs of compliance together with the punishment for non-compliance does not easily exceed the value of expected benefits of non-compliance. Thus, under this basic model, sufficient incentives to comply do not seem to be present.

### 5.3.4 Reaction of objects

The assumption that can be seen to underlie the preventive or deterrent theories is that the possibility of liability for damages caused by defective products is seen by the producers as a risk. To them, the law is something uncertain and unpredictable; a risk that has to be managed. Product liability risk is a category of legal risks that the vendors have to manage in any industry[976]. Enterprises as economic actors in general can be assumed to endeavour to analyse the risk they are subjected to in their line of business and their associated costs and to minimize them.

The expectation is that objects of regulation in the case of secure software development take preventive measures in order to shield against legal liability costs. The costs and benefits of liability

[976] This is emphasised by Janet R. Hunziker and Trevor O. Jones, *Product Liability axd Ixxovatiox*, p. 13, in an overview of more traditional industries like chemical, medical devices, automotive engineering, and general aviation engineering mainly in the U.S. This is clearly visible also in the management guides to computer security where legal issues, if considered at all, are considered briefly together with other operational risk issues. From the perspective of regulatory theory the fact that market actors see the law as a source of risk (it is something uncertain and unpredictable; a risk that has to be managed) is contrary to the traditional legal dogmatic view according to which the private law plays essentially a facilitative role and provides a map of market life by reflecting market practices, and calls into question these conceptualisation of (private) law as argued by Julia Black in Law and Regulation, p. 52-54. However, both views are typically present in dogmatic legal analysis to the degree both the preventive and reparative (compensatory) functions of private law are recognised.

management are assumed to be compared and an economically efficient level of precaution in the form of risk management be taken. Businesses are assumed to observe regulation, in this case the risk of legal liability, as a cost which needs to be minimised by cost-effective adjustments to their operations.

The assumptions of rationally self-interested and fully-informed decision-makers that underlie this economic model of prevention[977] can be closer to reality in the case of high-attention decision-makers like product manufacturers than it would be in the case of private individuals. They possess the opportunity, information, and incentive to think rationally and in advance about the kind and the level of risk they are prepared to create and the tradeoffs such choices involve. These actors can be expected to be more readily deterrable by liability regulation than ordinary private individuals acting as consumers who do not make decisions about potential liability in this way[978].

Preliminary empirical research on the assessment of liability risks, however, points to the direction that the intensity of the examination of the risks of liability depends on the extent, size, time-scale and general economic importance of the business transaction in question. According to the respondents from business associations and representatives from the legal professions to an EU wide questionnaire on contractual and non-contractual liability from the year 2001, exact and detailed liability risk analyses are undertaken only for very

---

[977] The influence mechanism of liability regulation, i.e., signalling to potential victims and potential injurers about the way they ought to behave, requires an assumption that those whose behaviour the law is seeking to affect are rational. They must be able to perceive that they can minimise their liability by taking precautionary actions of a particular kind and amount as Cooter and Ulen, *Law and Economics*, p. 330, explicate in relation to tort-liability system.

[978] Schuck, Tort Liability, p. 485, makes this argument in his assessment of the regulatory effectiveness of tort law. He refers to Latin H (1994) Good Warnings, Bad Products, and Cognitive Limitations, *University of California Los Angeles Law Review*, vol. 41. As research on risk has shown the majority of citizens do not evaluate hazards in the way technologically sophisticated analysts do. Instead of employing risk assessment, they rely on intuitive risk judgements, typically called 'risk perceptions' (Slovic, Perception of Risk, p. 220).

important transactions. With standard business transaction liability risks are examined scarcely.[979]

As noted by Baruch Fischhoff and John F. Merz, available empirical studies also suggest that when experts must rely on judgement, their thought processes often resemble those of laypeople[980]. They tend to exaggerate the extent of their knowledge and to oversimplify policy issues. Since the feedback from the court system about the requirements of the product liability law is severely delayed, lacks in detail, and is of limited usefulness due to the lawsuits arising from very small proportion of claims that are, in turn, a small subset of the injuries associated with a product, also the experts have real difficulties in making decisions that comply with the requirements of the law[981].

The precautionary measures vary according to the specific characteristics of the industry, the product and its delivery[982]. Risk management involves a variety of methods like quality assurance, information disclosure (e.g., safe use instructions), contracting, and insurance. Certain main strategies above the specific measures used to manage product liability risk can, however, be categorised. *Prevention*

---

[979] The results of the research have been reported by von Bar and Drobning in *Study on Property Law and Non-contractual Liability Law as they relate to Contract Law*, p. 413-414. According to von Bar and Drobning, insurance coverage replaces the examination of liability in many cases. However, particular attention always applies to questions of liability which are linked with the object of contract itself such as the liability for defects and consequential damage. True liability risk management is likely to be even scarcer in the software industry due to the methods for legal risk management having not yet reached any level of sophistication and having not been generally adopted in the IT industry, as argued by a London based solicitor Rachel Burnett in Legal Risk Management for the IT Industry, p. 61-67.

[980] Fischhoff and Merz, The Inconvenient Public, p. 169.

[981] Fischhoff and Merz, The Inconvenient Public, p. 171-172.

[982] The differences in the ways companies in different industries manage the risk of product liability is emphasised by Janet R. Hunziker and Trevor O. Jones, *Product Liability and Innovation*, p. 13, in relation to more traditional industries.

of risks is the primary method and includes measures like development of more secure products, obtaining of information about product risks and its dissemination, and product recall and repair. *Limitation and transfer* of liability risks includes, among other things, contractual measures and insurance.

As was noted above, vague and unpredictable product liability rules leave software vendors uncertain about whether a particular level of precaution will result in the court's finding them liable or not liable for product related damages. In such a situation, vendors can be assumed to take excessive precaution[983]. In the case of COTS software industry, the primary precaution seems to be avoidance or shift of liability due to the uncertainties, instead of developing more secure software[984]. The management of the risk of liability for defective and insecure software has largely focused on the limitation or transfer of risk by contracts. Liability disclaimers and exclusion clauses provide a relatively cheap and easy way to manage risks[985]. Improvement of

---

[983] Economic theory explains this in the following way. If the vendor overestimates the legal standard (simply makes a small random error in predicting the legal standard that could go either way) and since unnecessary precautions costs them less than liability, the vendor is given an incentive to take more precaution in order to create a margin of error within which they will not be liable. Note that this applies to a negligence rule. In relation to the rule of strict liability, court errors in computing damages distort precaution more than errors in setting standards. (Cooter and Ulen, *Law and Economics*, p.319-320)

[984] Preliminary empirical evidence from the research on the economic impact of the development risk defence as provided by the EC Product Liability Directive 85/374/EEC (Calderini et al., *Analysis of the Economic Impact of the Development Risk Clause as provided by Directive 85/374/EEC on Liability for Defective Products*, p. 46-47) demonstrates that when faced with high uncertainty and especially with possibly catastrophic consequences, only firms that have the capability to correctly evaluate the risks and to enact activities are likely to be efficiently incited to develop safer products. Other firms tend to over-react and to decide with respect to worst-case scenarios. The consequence could be the abandoning of the business altogether or, as seems to be the case in the software industry, the adopting of opportunistic behaviour (strict liability disclaimers) in order to reduce potential liability. This practice does not increase the security of software and leaves the customers with little or no protection.

[985] By disclaiming liability the risk of liability for software defects is managed in

the security of the software has not been of much concern[986], perhaps until recently, and insurances have not been available for long.

Despite the excessive reliance on contractual risk management techniques and the relative unavailability and lack of use of many of the other methods of liability risk management, in a regulatory analysis we still need to consider how the other possible risk management methods shape the influence mechanism of product liability rules on secure software development. Since we are concentrating on the capabilities of product liability rules to influence behaviour, and not their actual influence on behaviour, the *possible* reactions of the vendors need to be considered in order to see how the influence of product liability could be shaped by the reactions of the objects. This widened analysis is further justified by the increasing demands for regulation through product liability rules both in U.S. and Europe and the consequent increasing foresee-ability of liability related costs possibly leading vendors to reconsider the validity of liability disclaimers and exclusion clauses[987].

---

an international context with large customer base more easily and with lower initial costs than by other risk management measures; even more easily than by disclosing the vulnerability information. The argument made above that legal risk management has not been generally adopted in the IT industry also partially explains the heavy reliance on contractual risk management. It simply provides to most easy way to manage the liability risk.

[986] Liability disclaimers and exclusion clauses have been used successfully at least as a post-breach bargaining tool and a psychological barrier. Liability has not been foreseeable enough for the vendors to invest more in secure software development.

[987] Other risk management methods become relevant when the uncertainty about the validity of the contractual risk management increases. The different approaches among the legal systems of EU Member States to the validity of liability exclusion clauses, pointed out by Christian von Bar and Ulrich Drobning, *Study on Property Law and Non-contractual Liability Law as they relate to Contract Law*, p. 177-178, in their study on the problems and obstacles for the smooth running of the internal market resulting from differences in property law, non-contractual (tortious) liability law and contract law in most of the legal systems of the European Union, increase the uncertainty about the validity of liability disclaimers and exclusion clauses. At least the considerations of the extent to which liability limitation clauses are recognized and can be implemented in different jurisdictions, the risk of actual legal action (and the enthusiasm for litigation) together with costs of going to court all play an important role. The variety of approaches in turn raises the costs of

As pointed out above, the assumption of the regulators in product liability seems to be that the primary preventive measure is improvement of the products quality and security together with gathering and disseminating security related information[988]. The emphasis ought to be on these preventive measures to the degree the bugs are economically feasible to remove. The approach seems to be that the risk management methods that try to transfer or limit liability ought to be used only for the remaining risks. To the degree preventive measures like more secure development and information dissemination are taken due to the increase in potential liability costs, the assumption about the ability of product liability to influence secure software development is correct.

COTS software business relies heavily on patching to correct the bugs that are discovered from released products. Due to speed-to-market and short term cost reduction considerations this can happen even at the cost of initial secure software development, as has been argued above. This is a central shaper of the influence of product liability rules on secure software development. The focus of software development does not necessarily turn to adding security from the

---

ascertaining what the legal position of a specific country is, if it is ascertainable with anything like certainty. As Juha Pöyhönen, one of the leading contract law scholars in Finland, notes in Törkeä tuottamus ja vastuunrajoitusehdot, p. 97, the use of such liability risk management methods that are not legally certain, i.e., are either wholly or partly invalid disclaimers, can give an overly positive conception of the actual liability position for the vendor. Reliance on such uncertain terms can, at worst, lead to inefficient management of business risks.

[988] This is inline with the perspective shared by most scholars in the field of innovation economics and strategic management, who agree that the way in which companies change and adapt to the environment consists in modifying and introducing new procedures and norms. According to this perspective, firms' operations are essentially based on routines, norms and procedures that may be formal or informal and tacit. As noted by Mario Calderini, Marco Cantamessa and Alessandro Palmigiano, *Analysis of the Economic Impact of the Development Risk Clause as provided by Directive 85/374/EEC on Liability for Defective Products*, p. 43, in their analysis about the economic impact of the possible removal of the development risk defence provided by the European Product Liability Directive 85/374/EEC, according to this view, companies will deal with a change in the legal liability regime by adapting their internal and company-specific product development process to the extent that it is possible or economically viable.

start, instead the practice of release-and-patch could continue. This is in case the companies seriously turn to quality and security assurance as risk management methods, instead of strict liability disclaimers. As has been argued above, contractual risk management has dominated for long in the case of software security risks.

*Effects on innovative activity*. Product liability rules not only influence the firms decisions to use secure software development measures, but they can also affect their willingness to innovate. The effects on innovative activity are, however, very difficult to analyse and subject to many variables[989]. Only certain basic arguments can be made without a detailed industry specific analysis that is not available at this stage.

One of the strategies for firms to cope with increased uncertainty and risk in a stricter liability regime, is a reduction of their innovative output. However, the empirical evidence from more traditional goods provided by Kip Viscusi and Michael Moore supports this approach only partially[990]. Higher product liability costs provide incentives for product safety improvements and increase product R&D intensity. However they do it only initially, and at high levels of liability costs, liability reduces innovative activity.

As argued by Kip Viscusi and Michael Moore at a theoretical level, higher liability costs will increase product innovations directly related

---

[989] As pointed out by Calderini et al. in *Analysis of the Economic Impact of the Development Risk Clause as provided by Directive 85/374/EEC on Liability for Defective Products*, p. 133, the effect on innovative activity is shaped by many different industry specific characteristics, such as market structure, the pace of innovation, product architecture and the product life cycle. In the terms of Janet R. Hunziker and Trevor O. Jones in *Product Liability and Innovation*, p. 14 "[t]he synergy between the product liability system and innovation takes place in a particular social, legal, and regulatory environment. The peculiar nature of that synergy is influenced by trends in such things as the availability of insurance, public attitudes toward risk, the body of statutory law that governs products and processes, and how technological questions are dealt with by the law."

[990] Viscusi and Moore, Product Liability, Research and Development, and Innovation, p. 182.

to safety improvements and also those that introduce new technologies if they decrease the costs of providing safety[991]. Thus, not only the use of secure software development tools and methods would be positively affected by increased liability, but also the security improving innovations would be increased. However, innovations that do not lower the marginal costs of providing safety will be depressed. In the case of extreme liability costs, product novelty will be eliminated altogether as the firm selects the no-risk corner solution.[992]

It has been argued that if the software vendors would face full liability of the design defect types of software vulnerabilities, the costs could be so big as to make not only the profitability of the product but possibly even the firm itself precarious. It is true that the highly innovative software firms that introduce new products with uncertain implications for safety and security also run substantial liability risk. When insurance is not available, and is not likely to be in the near future for innovation fuelling SMEs due to the high costs, as will be discussed below, the costs of stricter liability does seem unbearable for many companies. In such a situation, the effects of liability rules could be devastating.

In the case of COTS software, however, this would be unlikely. Even though software vendors have feared the theoretical possibility for full liability for any consequences (direct of indirect) of bugs and malfunctions of the delivered software and issued exclusion clauses into license agreement attached to the software, their actual extent does not seem overwhelming. The many excuses and defences provided by product liability rules lower the true liability risk. As the above analysis of the substance of product liability rules showed, under a deeper analysis the risk of liability does not seem unbearable at all for the software vendors in many countries; even if it would be applied

---

[991] Viscusi and Moore, Product Liability, Research and Development, and Innovation, p. 167. The empirical data used by Viscusi and Moore did not enable them to distinguish the effects of product liability on safety-related research and development expenditures and on the development of new varieties of products (Idem., p. 182).

[992] Viscusi and Moore, Product Liability, Research and Development, and Innovation, p. 167.

in its current form. But the high costs of developing more secure software combined with increased liability and insurance costs *might* negatively influence SMEs' willingness to innovate. Research into these issues is needed.

*Use of insurance.* Possible increase in the probability of liability for insecure software is also likely to lead to increased demand for insurance or financial coverage. However, cyber-insurance markets are just emerging[993]. The costs of insecurity of such a relatively new type of a complex product such as software have been unknown and insurance has not been offered for long[994]. The possibility for insurers to rely on measures of predictability to forecast probable risk and set prices have been enabled only recently when scant historical and actuarial data about the risks of product liability in the case of insecure software is being gathered. For the same reasons, prices have remained high[995]. They are relatively beyond reach of small and medium-sized

---

[993] This has been pointed out in 2003 by Christian Egeskov and Jan Christensen in Behovet for forsikring af softwareproducentens mulige erstatningansvar, p. 51, in relation to the U.S. and the Danish markets, and in 2005 by Jay P. Kesan, Ruperto P. Majuca and William J. Yurcik in The Economic Case for Cyberinsurance, p. 26 in relation to the U.S. insurance markets. This situation has not changed much since 1977 when Peter Seipel in *Computing Law*, p. 87, pointed out that "[l]iability insurance for programming consultants, software houses, and similar categories is, however, available only to a rather limited extent".

[994] As Connolly, Insurance: The Liability Messenger, p. 132, explains in discussing the reasons why the insurance system sometimes fails: "For new products, the data do not yet exist. And for complicated products, the available data do not provide a sufficiently strong basis for making credible predictions".

[995] Premiums for cyber-insurance that cover several areas including losses arising from DoS attacks, e-business interruption, electronic theft of sensitive information etc, in addition to product liability suites, have been reported to range from $5,000 to $60,000 per $1 million of coverage (Kesan et al., The Economic Case for Cyberinsurance, p. 29). The prices can expected to continue to remain high since the time and stability to develop statistical data for actuarial tables is just not present in the case of rapidly moving software security environment. Software products are in constant change and as a consequence, flaws in software change dynamically and new attacks are released daily. Future risks are unknown also because both crackers and

companies. Still, vendors' insurance is certainly one way to provide victims with compensation while also allowing companies to afford high risk activities such as innovation and product development.

When considering the way the increased demand for insurance shapes the effects of liability, despite the above problems of insurance markets in dealing with losses arising from insecure software, we cannot bypass the moral hazard problem. As explained in microeconomic theory, moral hazard arises when the behaviour of the insured changes after the purchase of insurance so that the probability of loss or the size of the loss increases[996]. Insurance policies are considered under this theory to affect firms' incentive in taking adequate care in product safety and security. However, the effect of insurance policies on product security is unclear and multi-faceted[997].

The influence of the insurance system for the management of corporate liabilities is best understood, as explained by Dennis R. Connolly, by considering the system as a message-bearer[998]. Insurers send a message to insurance buyers whenever they set premium rates and establish the terms and conditions of coverage. Already when the insurance companies are in the position of monitoring injurer's procedures, the injurer will have an incentive to take efficient care. However, since information asymmetries almost entirely prevent efficient monitoring of software development practices, the insurer's control over company activities might not be effective in improving the incentives for secure software development. As Mario Calderini, Marco Cantamessa and Alessandro Palmigiano explain, insurers may try to resolve information asymmetries by acquiring information and

---

protective technology is getting better. (Kesan et al, The Economic Case for Cyberinsurance, p. 29)

[996] Cooter and Ulen, *Law and Economics*, p. 50.

[997] Calderini et al., *Analysis of the Economic Impact of the Development Risk Clause as provided by Directive 85/374/EEC on Liability for Defective Products*, p. 70.

[998] Connolly, Insurance: The Liability Messenger, p. 131-132. In the terms of this study, they regulate the behaviour of the vendors with a market-based instrument. Insurance would deserve to be studied as a type of regulatory instrument in its own right. Only its relations to the liability rules are briefly considered in here.

setting premium-performance schemes[999].

Basically, the insurance company can require software vendors to undertake secure software development tools and practices and to employ standards and development methods where available. Premiums can also be tied to claims histories where available or proactively to the insured firm's investment in secure software development. Naturally this is not without problems for example in the establishment of correct standards and in their private enforcement by the insurance companies. However, this should, at least in principle, give sufficient incentives for the software vendors to improve the security of their products contrary to the moral hazard argument[1000].

The turn to insurance industry for help in the managing product liability risks might not only lead to more secure software but also to a level of security that is socially efficient. In the most positive evaluations such as that of Jay P. Kesan, Ruperto P. Majuca and William J. Yurcik the pooling of information by the insurance companies and their superior expertise in assigning proper prices to risk and in developing safety standards is seen as a way to set achieve the socially efficient level of care[1001]. This might be an overly optimistic view of the capabilities of the insurance companies, but it is true that they can at least help in setting regulatory standards and make it easier to find the socially optimal level of care.

---

[999] Calderini et al. in *Analysis of the Economic Impact of the Development Risk Clause as provided by Directive 85/374/EEC on Liability for Defective Products*, p. 70. Most common methods to minimize moral hazard are coinsurance and deductibles. Under coinsurance the insured party shoulders a fixed percentage of her loss and under a deductible plan, a fixed monetary amount of the loss, with the insurer paying for all losses above that amount. Some insurance companies also offer reductions in premiums for certain easily established acts that reduce claims. (Cooter and Ulen, *Law and Economics*, p. 51)

[1000] Note that this does not have to restrict to external influencing by the insurance companies. Info *on* the availability of premium reductive tools and development methods can even change the preferences (intrinsic predispositions) of vendors in relation to secure software (desire for security instead of pure features and functionalities).

[1001] Kesan et al., The Economic Case for Cyberinsurance, p. 19. This is the situation where the cost of a little more precaution (marginal cost) equals the resulting reduction in the expected cost of harm (marginal benefit) (Cooter and Ulen, *Law and Economics*, p. 301.

# 6 Conclusions

As a partial answer to the question posed by Henrik Kaspersen about the issues that should be enhanced in information security, this study postulates that secure software is a central element. Secure software development is one of the most crucial issues in information security to be enhanced. It ought to be an important public policy if we wish to achieve a more secure network society and to maintain trust for information products and systems in commerce.

Of course, the entire blame for information insecurity cannot be placed on the developers of vulnerable software. However, without such a high amount of exploitable defects in software the attacks by opportunistic criminals and the accidents due to the behaviour of computer users would result in far less damage. A significant part of the problem can be traced to poor software quality and security.[1002]

This is so especially now when we are moving to a mobile world where services utilising Internet connections are more and more used with portable devices. This trend together with the wider digital convergence increases the complexity of systems and at the same time brings new sources of information security threats to the mobile networks, devices and services used therein and to their users. In addition to a great number of conventional threats from the personal

---

[1002] As the Canadian law professor Jennifer Chandler also correctly emphasises in *Improving Software Security*, p. 3, the problems of information security are probably best tackled on many fronts simultaneously, including efforts to enforce cyber crime laws, measures by internet service providers to help secure the network, improvements in the security consciousness of end-users and, concerted efforts to improve software security. However, secure software development needs to be placed among the most important issues to be tackled with. Despite the increasing interest, it still remains neglected when compared to the issues of network security and computer crime.

computing area that concern also mobile devices[1003], a variety of totally new set of security risks come along with wireless networks[1004].

In a situation where the level of information security management, together with know-how about information security issues, varies between different stakeholders in such a rapidly evolving field as mobile communication, the reliance on reactive protection measures is no longer sufficient. Especially not when mobile devices are used in ever more critical situations (e.g., mobile devices in professional business, radio frequency identification RFID in passports), and the use of the devices and services, together with their development, is no longer controlled by a single entity[1005]. The great number of software versions in mobile devices together with the problems in keeping them updated[1006], unsatisfactory maintenance of software by users (e.g., updating anti-virus software and backup of data) and the more risk prone open programming interface of the new mobile operating systems (e.g., use of a general purpose programming languages like Java) emphasise the importance of the security of

---

[1003] In a recent research of the threats to information security in the use of mobile devices, services and networks done for the Development Programme on Trust and Information Security in Electronic Services (LUOTI) of the Finnish Ministry of Transport and Communications, Ahonen et al., *Information Security Threats and Solutions in the Mobile World*, p. 41-44, report such threats as Denial of Service (DoS) attacks, virus infections and spam.

[1004] Basie von Solms and Emil Marais, From Secure Wired Networks to Secure Wireless Networks – What Are the Extra Risks?, p. 634-635, identify such risks as lack of proper risk analysis due to the easy installation of wireless networks, eavesdropping, and ad-hoc peer-to-peer networks even without a need for a wireless access point.

[1005] As pointed out by Ahonen et al. in *Information Security Threats and Solutions in the Mobile World*, p. 23, when the network infrastructures are converging towards IP-based solutions, the control and responsibility of the information security management is becoming fragmented.

[1006] The possibilities for automated updating of the software in mobile devices is not only currently under developed, but also includes a lot of difficulties, for example, in maintaining the integrity of the updates and verifying the sources of updates.

software used in them[1007]. Security needs to be developed into the systems form the beginning. Secure software development becomes even a more crucial issue when software in a mobile device is being used in an ever more ubiquitous manner. It is no longer sufficient to rely on the software patching system alone (and the independent security researchers who detect and publicize flaws) or the secondary market in security software such as firewalls.

The argument has been that the state of the art in secure software development in COTS markets is not sufficient for the needs of the converging information infrastructure and the devices and services used therein, where software products are increasingly being used as components. The level of security just is not enough for electronic services such as e-commerce or e-government. A central element for the success of electronic services, i.e., trust in to the systems in use, needs less vulnerability. Even more central the problem is for those critical infrastructural systems, such as communications and electronic networks that use COTS components.

The traditional ways in which the state of the art in product safety generally improves in an industry simply are not sufficient. For long, COTS software industry have not dedicated sufficient amount of resources to improve the security of their software. The care exercised during the product development process, e.g., in requirements engineering and product testing, are no longer sufficient and the interest in the root causes of potential harm caused by vulnerable software is not high.

This has been mainly due to the immaturity of software engineering as a discipline and consequent tradition of accepting defects in software. Software security has not improved largely also due to the lack of both demand from the customer side and of other incentives for secure software development in the information economic environment characterised by network effects and lock-in, where market-driven software development occurs.

---

[1007] These problems in the mobile world have been explicated, e.g., by Ahonen et al. in *Information Security Threats and Solutions in the Mobile World*, p. 44-45.

Not even the other mechanism in which state of art in product safety generally improves have been sufficient to improve the security of COTS software to the socially sufficient level, i.e., companies' reaction to the information about the failures of their products on the marketplace. Even though users are increasingly being enlisted to help in finding errors in software products by making early releases (beta versions) available, the problem is that this does not enhance the security of the software products. Mainly due to the lack of skills and motivation for conducting sophisticated security-specific testing needed to find software defects causing security related vulnerabilities, no amount of beta testing by customers at large will cover a sufficient amount of bugs. As a result, an insufficient amount the security related problems are reported to the developers. In addition, when patches are released they often only fix the symptom of the problem thus leaving the cause unaddressed, they often introduce new problems due to them also being rushed out as a result of market pressures, and often also go unapplied or are otherwise ineffective.

As the analysis made in this study suggests, it seems that industry custom for mass-market software leaves significant room for improvement[1008]. Current practices in the market-driven software development do not necessarily reflect a reasonable level of care[1009]. The costs of improvements in secure software development do not yet equal the resulting reduction in the expected cost of harm when considering all the societal costs. This means that it is possible to

---

[1008] Canadian law professor Jennifer Chandler also emphasises this in her recent article *Improving Software Security*, p. 22.

[1009] Current market practice for developing COTS software seems to follow the 'release and patch' mode of selling software in the mass-market software industry. Everyone is thought to understand and expect that complex modern software will contain bugs. As Jennifer Chandler in *Improving Software Security*, p. 21-22 also notes, this customary practice can be a persuasive evidence of what is considered to be the reasonable standard of care in the software industry. Conformity with the custom usually is an indication of non-negligent behaviour. However, compliance with custom is only a helpful evidence of the exercise of reasonable care. It is not conclusive evidence. A court may conclude that a custom is unreasonable and legislative activity can change the required standard of care.

increase the level of security in software, especially the amount of security related defects in it, without losing the immense utility its use have brought and can bring[1010].

The sadder end result of this study is that even if the motivation of the market actors would increase, e.g., customers would start to demand secure software as we are currently witnessing, the level of security will not rise to a level sufficient for the needs of the society and its information infrastructure. As the considerations of such market failures as externalities and inadequacy or asymmetry of information suggested, the private motivation to enhance security might not be sufficient to provide the amount of security that the society as a whole needs for its information infrastructure. If the market actors are left to themselves to enhance the technological and methodological tools for software development, security of software might not rise to a sufficient level from the societal point of view, i.e., from the point of view of the security of the information infrastructure that uses the software developed by the COTS software industry[1011].

If you accept my argument about the reasons for the lack of security in software, the answers to the questions made by Henrik Kaspersen about the factors that should be enhanced and the role of regulation becomes more evident. The main problems that regulation ought to try to solve in secure software development are negative externalities and inadequacies or asymmetries in the distribution of security related information[1012]. Negative externalities

---

[1010] Since greater security can be achieved, as has been argued several times above, the relevant inquiry is whether the costs of improving security is more or less than the value of the improved security (Chandler, *Improving Software Security* , p. 20-21). Detailed calculations have not been made, but as the lack of consideration for social costs by the vendors testifies, there is ample room for improvement.

[1011] Without downplaying the importance of the enhancement of the methods and tools for secure software and information systems development, since they are the means by which the security of software can improve, it must be acknowledged that they are not improving without sufficient incentives.

[1012] Even though the market failure considerations do not alone justify the need

are the costs spilled over to third parties due to the abuse of vulnerabilities in common software product, such as system recovery and upgrade, loosed business time, costs of dealing with the breaches etc. Lack of adequate information about the quality aspects of security in software leads to failure in the production and sale of information in relation to the quality of software and its security features. The informational asymmetry favouring the software vendor prevents the underlying software products market from working properly and may lead to more secure software products being turned out from the market.

Two regulatory tools of vulnerability reporting and software product liability were chosen for a deeper analysis in this study as likely candidates for solving these problems. Both of them involve theoretical potential for inciting improvements to the current secure software development practices. They try to influence the above mentioned mechanisms by which the state of the art in product safety generally improves. Potential software product liability tries to increase the motivation to dedicate resources to secure software development and vulnerability disclosure mainly tries to improve the efficiency of the feedback mechanism.[1013]

As regulatory tools, vulnerability reporting and software product liability face several problems as has been argued above. Their influence mechanisms are shaped by so many negatively shaping factors that the practical effects are largely diluted. However, they are not without capacity to influence. It is just that these regulatory

---

for regulation, since they are only one of the technical justifications for regulation, they do show essential points to be regulated and tell about the importance of regulation. They tell about the places where regulation is needed and suggest what it can do. Welfare economic considerations are not the sole justifications for regulation, but they are essential in depicting what to regulate and how.

[1013] Even though the concentration in this study is on the secure software development part of information security, it needs to be stressed that the regulatory tools analysed in this study influence, in addition to the behaviour of software vendors and developers, also the attitudes of users and likely also the users of COTS components.

instruments would have to be altered in order to be able to influence secure software development in practice.

It needs to be accentuated again that while considering ways to enhance the efficacy of the two regulatory instruments analysed in this study, the purpose is not to give direct advice in public policy. The effort is to analyse different means to enhance certain public policy, i.e., that of secure software development. Full evaluation of the feasibility and desirability of different regulatory solutions for the problems of developing secure software would require information much more widely than just of the regulatory capabilities of instruments. Issues like the efficiency, equity, manageability, legitimacy and political feasibility of the specific instruments or their mixes would, at least, be relevant. As the research also showed, several other instruments would also have to be considered in similar vein.[1014]

## 6.1 Improving the influencing capacity of software product liability rules

*Justifications for change.* The pressure to make product liability for defects a serious threat for COTS software vendors has already increased[1015]. It is no longer just the practitioners that are presenting demands for

---

[1014] Similar to the role assigned for regulatory impact analysis in general in a 1997 collection of best practices in OECD countries (OECD, *Regulatory Impact Analysis*, p. 7), the analysis of the regulatory capability of specific instruments made in this study is not a sufficient basis for regulatory decision-making. At best, it is only a guide to improve the quality of political and administrative decision-making, while also serving important democratic values of openness, public involvement and accountability.

[1015] Product liability rules, are understood in this study to cover both laws deriving from the EC Product Liability Directive (Council Directive 85/374/EEC of 25 July 1985 on the approximation of laws, regulations and administrative provision in the Member States concerning liability for defective products, OJ L 210, 7.8.1985 pp. 29-33, amended by directive 1999/34/EC OJ L 141, 4.6.1999, pp. 20-21) and general doctrines for liability for damages under contract law or tort law (delictual liability).

software product liability; also the political discussion is rising. The demands for increased liability are likely to continue.

For long, software industry has shielded itself from liability due to the immaturity of the business and the methods of secure software development. However, the argument that the state-of-the-art in software development does not enable vendors to avoid security related vulnerabilities or at least discover and correct defects that cause them is no longer valid. Even though software development practices traditionally have not acknowledged security related vulnerabilities and cost-effective ways to avoid them have been scarce, with increased interest in security related vulnerabilities such methods have evolved.

As has already been argued, despite the current inadequacy of the methods and tools for secure software development there are well-known technical and procedural ways to prevent at least the widely known vulnerabilities; the state of the art no longer prevent vendors from doing this. It is more of a result of the business environment; first-to-market competition in the network economic environment motivates to release early and often which tends to override the time consuming and expensive security considerations.

Thus, the release of unfinished products should no longer be excused from liability under the state of the art –defence, at least not when information security is threatened. When vendors have cost-effective means to discover and correct vulnerabilities, the beta-release-as-a-final-version -type of practice, where the release is followed by extensive patching and corrections in new releases, should be allowed only to a degree. It no longer is an adequate justification to ignore the expectation of security that software product users have.

Here we come to the degree to which software vendors are liable under product liability rules[1016]. The standard that the default rules

---

[1016] Bear in mind that in this study, product liability rules are understood to cover both laws deriving from the EC Product Liability Directive (Council Directive 85/374/EEC of 25 July 1985 on the approximation of laws, regulations and administrative provision in the Member States concerning liability for defective products, OJ L 210, 7.8.1985 pp. 29-33, amended by directive 1999/34/EC OJ L 141, 4.6.1999, pp. 20-21) and general doctrines for liability for damages under contract law or tort law (delictual liability).

on defectiveness (conformity) and consumer product liability rules set is dependent on the level of quality and security their customers can reasonably expect.

As has been argued above[1017], customary practice in secure software product development leaves significant room for improvement. Current practice in secure software development is not optimal. Since improvements seem to be cost-justified, current practice does not reflect a reasonable level of care in software development. Neither does it reflect the level of expectation users' have on the security of the software products that they buy.

Many issues have traditionally spoken against the expectation of security on the customers' side. It has been a widely repeated fact that defect free software is an illusion and no customer can expect any complex software products to be flawless. In addition, customers have been accepting towards vulnerabilities (even security related) resulting from the defects as a trade-of for new innovative features.

Currently, however, industrial buyers (as system developers) are increasingly recognising their high self-interest in the security related vulnerabilities. They are increasingly realising that vulnerabilities found in the components of their systems not only will cause damages when the risk materialises[1018], but will also hamper their sales of the system or damage their public image when the existence of the vulnerability also in their system (that uses the vulnerable component) becomes widely known. They thus have legitimate expectations for the security of the components they use. Even the negative effect of vulnerabilities (especially when they are widely discussed in the media, as they

---

[1017] See chapter 2.

[1018] The direct and indirect damages might be substantial, e.g., in the case of an e-commerce site being shut down due to a DDoS attack via a vulnerability in a component of the system. Cambell et al. in The Economic Cost of Publicly Announced Information Security Breaches, p. 431-448, provide empirical evidence also of significant negative stock market reactions for published information security breaches involving unauthorised access to confidential data (especially if personal data is accessed).

currently are) on the overall trust of customer in the line of business at least ought to speak in favour of the expectation of security.[1019]

Also private individuals in general at least ought to have good reasons to expect security from software products. Even though their self-interest in claiming damages or resorting to other remedies is low due to the direct losses for the individual end-user in many cases being close to insignificant (such as in cases of her computer being used as an unconscious host for DDoS attacks), as a group they also suffer significant losses due to vulnerabilities in their systems being abused. In addition, the costs that users have to bear in trying to keep up with the increasing pace of fixing bugs or replacing faulty software are going to be high in the long run.

As the software industry has matured and the vulnerability issues have been widely discussed in the media, individual interest in the level of security their software have also been increasing. Especially in cases where security is the issue (such as firewalls, encryption products, virus scanners etc.) the expectation of lack of security related vulnerabilities can easily be justified.

What this means is that the customer expectation currently is software that does not pose unreasonable risks. Customers still have to extend reasonable tolerance when defects manifest themselves, but it is clear that at least the expectation of the widely known security related bugs not existing is justified[1020].

---

[1019] Note that for the same reasons industrial buyers are more likely to negotiate on these issues expressly in software contracts and have the leverage to demand changes to the standard form contracts typically used in COTS software business. Even though such practice is still rather rare, it will reduce the importance of the default rules that depend on the expectations of the users in general as the practice generalises. However, bear in mind that the investment in security by one organisation in a networked environment creates sizeable external benefits (positive externalities) of which the investor does not get compensations and, similarly, the breach of security brings negative externalities. The incentives to invest resources are lower than the societal optimum when these costs and benefits are not considered.

[1020] As Chandler, *Improving Software Security*, p. 28, puts it in relation to tort liability: "In the end, perhaps the strongest arguments about the reasonable standard of care in software development can be made about common

Under the law, the final level of quality and security that the buyer has a right to expect is determined by the contracts and complementary marketing information[1021] and is dependent on the level of security provided in the market and on the customary market practices. However, acknowledgement of the general expectation of security that the customers have on software products ought to be already enough, however, to show that the reasonable expectation of quality and security is higher than the level of security provided today. The requirement for changes in the level of security provided can be justified on the basis that customers expect higher quality and security from the purchased software than is currently being provided. The justified level of expectation has increased from what it has been.

If the argument of previous customer acceptance of insecurity is taken seriously, the time when this tolerance vanished has relevance to the determination of the defectiveness of the software. As the year 2000 problem taught, the time when the state of the art could have enabled the vendors to develop secure software in a cost-effective way shows the critical timeframe after which the customer expectations for secure software are justified. However, if the time when the customer expectations shifted is different, then this is dominant. Establishing such a timeframe is a question to be solved

---

implementation errors. It seems reasonable that, at a minimum, the presence of well-known, easily detectable, and easily remediable flaws should be considered negligent."

[1021] They are the main norm source in the interpretation of contractual situations. The business practice is a secondary source, and the default rules in sales of contracts regulations apply only if these legal sources are silent or unclear. Information given of a software product is important also in relation to the determination of the defectiveness of product under consumer product liability rules. Under consumer product liability rules, such as the EC Product Liability Directive (1985/374/EC) and its Finnish implementation (Product Liability Act 694/1990) the specification of the required security standard is dependent on a multitude of factors. Relevant factors can include, in addition to the expected usage of the products and the information provided in combination with the product and its use, issues like the expectations of the users, the level of security in the relevant markets, applicable governmental orders and relevant standards.

in practical cases and would require further analysis of the technical possibilities and practices in software business. No overall timeframe is likely to be established that would concern all types of software or even all types of defects.

*Means for change.* Product liability rules, as understood in this study to cover both laws deriving from the EC Product Liability Directive (85/374/EEC)[1022] and general doctrines for liability for damages under contract law or tort law (delictual liability), involve theoretical potential for inciting improvements to the current secure software development practices as is shown above in the analysis of the influence mechanism[1023]. This investigation of the efficacy of private law liability as a form of regulation of product quality and security suggests that liability rules possess considerable capacities to influence secure software development.

The default defectiveness rules in contract and product liability together with marketing rules and tort law rules seem to provide incentives both for developing more secure software and disclosing vulnerability information to the customers. There are limitations and vagueness's but by making the assumption that these rules are applied and effective they seem to provide effects that are overall beneficial for secure software development. The implicit argument is that product liability rules ought to be applied to COTS software.

The influence mechanism of product liability rules is shaped by many intervening factors; even to the degree that their effect on secure software development might be diluted altogether. This is why product liability system would have to be altered in order to be more efficacious in its regulation. Without repeating the structural weaknesses of private law system of regulation in standard setting,

---

[1022] Council Directive 85/374/EEC of 25 July 1985 on the approximation of laws, regulations and administrative provision in the Member States concerning liability for defective products, OJ L 210, 7.8.1985 pp. 29-33, amended by directive 1999/34/EC OJ L 141, 4.6.1999, pp. 20-21

[1023] See heading 5.2 above.

monitoring and enforcement discussed above[1024], we can say that the one of the most important factors shaping the effects of product liability rules on secure software development, which is at the same time characteristic for contracting in the COTS software market, is the current contractual practice in the mass-market software industry of extensively excluding and disclaiming liability.

Software vendors have feared the theoretical possibility for full liability for any consequences (direct of indirect) of bugs and malfunctions of the delivered software and for this reason issued exclusion clauses into license agreement attached to the software. However, the actual extent of product liability does not seem overwhelming in a deeper analysis. The many excuses and defences provided by product liability rules seem to lower the true liability risk. As the above analysis of the substance of product liability rules showed, under a deeper analysis the risk of liability does not seem unbearable at all for the software vendors in many countries; even if it would be applied in its current form.

Neither are the arguments from the industry, e.g., that innovation would be hampered if liability costs are increased, convincing under the evidence from more traditional industries. However, due to the specificities of the software industry more research is needed in order to validate this claim. Only if it can be shown that software vendors can manage even under renewed liability rules (e.g., when the uncertainty about the liability risks diminish, and this is a work for the regulators), are changes in the vendor's ability to shield herself from liability for defective software feasible. Even then changes might be feasible only to the degree vendors can still make profit under the new liability rules, i.e., the possible legal actions would not cripple them)[1025]. It is not wise to alter the industry practice without context

---

[1024] See heading 5.3 above.

[1025] Suppliers may validly argue that increased liability will drive up the costs of software. Customers should realise, however, that the increase in the upfront cost may in the long run be less than they are currently spending on attempting to fix bugs or to replace faulty software. See chapter 2 above. The same argument is made also by Dominic Callaghan and Carol O'Sullivan in Who

specific knowledge since the production of software has for long been made possible with the liability disclaimers. Still, the social benefits from increased liability are large enough to put the regulatory approach in a favourable light even without such context specific knowledge. In addition, a change in the vendors' liability for defective software is likely as the industry matures.

However acceptable the shifting of liability to the users might be, which still is largely a matter of dispute especially in consumer relations, the vendor incentives to develop more secure software and to disclose vulnerabilities do not seem to be sufficient at the presence of extensive liability disclaimers and exclusion clauses. When even the basic defectiveness rules on fitness for purpose or use are disclaimed and the force of mandatory rules is circumvented by licensing the software instead of selling it, the incentives for secure software development and disclosure of vulnerability information are widely diluted in the software industry.

The incentives provided by product liability rules are close to non-existent in relation to secure software development as long as standard form liability disclaimers are successfully used (i.e., are uphold in courts, are given legislative approval, continue to be untested in courts, or invalidating court cases do not receive sufficient media attention) and the vendors continue to rely on them extensively[1026]. Substantial

---

Should Bear the Cost of Software Bugs?, p. 60. One thing that requires further research is the likelihood in the increase of piracy if the upfront costs would increase, together with the possible increase in the market share of open-source software.

[1026] The consequences of using unfair terms that shift the burden of risks by externalising the costs in question, as is the case in liability disclaimers and exclusion clauses in EULAs', have been put forward in a short manner by the European Commission in its Report on the Implementation of Council Directive 93/13/EEC of 5 April 1993 on Unfair Terms in Consumer Contracts, COM(2000)248 final, p. 13, in relation to unfair contract terms in general: "firstly, the prices of products and services do not reflect true costs, creating distortions to competition in favour of less efficient firms and leading to lower quality products and services; secondly, the costs incurred by society are higher, because the risks and obligations are borne by persons other than those who could bear them most efficiently from the economic viewpoint".

changes in contractual practice (especially EULA's) would be required for the liability rules to have a real effect on secure software development.

The most desirable path to changes in software contracting practices towards more balanced approaches to liability for defects, that could turn the incentives for software development towards security, is for the market actors themselves to change the contractual practices. Currently, most contracts are still silent on conditions related to the security of the software to be delivered. This is so despite certain firms experimenting with the use of contractual clauses setting the software vendor liable for security breaches connected to its software[1027].

In certain sectors, however, such as in web applications, legal and especially contractual work is enhancing in terms of security. The OWASP Legal Project provides materials that help participants in the software market to discuss security and to capture the important legal aspects of secure software[1028]. The purpose of the OWASP Secure Software Development Contract Annex is to enable software developers and their clients to negotiate and capture important contractual terms and conditions related to the security of the software to be developed or delivered. Instead of taking security issues to court, cases involving weak application security are settled by informal means. The set of negotiable terms that it spells out are intended to be appended to software development contracts. Also organisations,

---

[1027] The trend been reported in the U.S., e.g., by Dennis Fisher in an article published in the eWeek on April 15, 2002, Contracts Getting Tough on Security, available at http://www.eweek.com/article2/0,1895,1658531,00.asp [29.3.2006].

[1028] The OWASP Secure Software Development Contract Annex is developed by the Open Web Application Security Project (OWASP), an open source community accompanied with a not-for-profit charitable organisation (OWASP Foundation). One of the stimuli for its development is the silence of contract on conditions related to the security of the software to be delivered, as pointed out in the "Introduction" to the OWASP Secure Software Development Contract Annex, available at http://www.owasp.org/documentation/legal.html [5.1.2006].

like the U.S. Federal Trade Commision, are enforcing rules that require especially those processing personal data to take security measures[1029].

The OWASP Secure Software Development Contract Annex is an example of a turn towards more balanced approach to contracting in secure software development. Developers are assumed to take the risk for problems that were covered in the requirements accepted by the customer as part or the provided security documentation or should be covered by reasonable testing efforts. Remediation of other security problems, called "novel" security issues, i.e., those not covered by the security requirements or outside the reasonable scope of security testing, is to be paid for by the customer.[1030]

As a national example, the Finnish Government Information Security Management Board (VAHTI) has produced a set of checklist for the use of government procurers in 2001[1031]. The guideline

---

[1029] Dave Stampley, in Privacy Compliance Enforcement, Part I, provides a list of privacy enforcement actions related to application security from the U.S. between 2000 and 2005.

[1030] Note that this standard comes close to the one provided in the EC Product Liability Directive 85/374/EEC with the possibility for the state-of-the-art –defence (Article 7 (e)) where there is no liability for defects that could not be discovered under the state of the art at the time when the product was put into circulation. Note that there is significant controversy over the state-of-the-art –defense in the EU. Examples of its successful use in courts in any EU country are difficult to find. However, the insurers and those producing goods clearly continue to regard it as important, whereas representatives of consumers suggested it be abolished (Meltzer, Freeman and Thomson (2003, p. vi). According to a recent EU study on the economic impact of this so called "development risk clause" provided by the EC Product Liability Directive 85/374/EEC Article 7(e) (Calderini, Cantamessa and Palmigiano 2003, p. 3), it has played a crucial role in finding the right balance between consumer protection and innovation in Europe, well beyond its limited use in courts.

[1031] Valtionhallinnon tietoturvallisuuden johtoryhmä, Valtion tietotekniikka-hankintojen tietoturvallisuuden tarkistuslista, VAHTI 6/2001. Only in Finnish. This is an important complement to the standard conditions of IT procurement contract in the Finnish Government (VYSE 1998), where information security is considered only in relation to the contracting for hosting services and maintenance. It can also be used in relation the multilaterally drafted Finnish contract template IT2000 Terms and Conditions for IT Procurement in Finland.

suggests best practices for different types of procurements, including prefabricated software products. The importance of clear contracts and the need to negotiate security terms in the very beginning of the procurement process is underlined. Central information security issues and requirements ought to be in the know and be clearly presented already when inviting for tenders.

The promise of such cooperative implementation measures is that the standards for secure software development can not only be drafted by the market participants in a much more cost-efficient way than any court could, but they might also be internalised better in the enterprises' decision-making concerning secure software development. This can be expected to lead to wider acceptance of the rules because the juxtaposition and conflicts between the parties are not emphasised.

The private law system of regulation could also use these contracting guidelines to help in reading basic security requirements into contracts that are silent on the issue[1032]. The contracting guidelines and checklists are good starting points in defining such reasonable requirements that both parties would have agreed to if the topic had been discussed. Naturally, it is much more beneficial for the software industry if the buyer and the developer work out the real security requirements for the project and capture them in the contract, rather than rely on what a court might read into the contracts when no specific requirements have been negotiated.

Unfortunately, such voluntary approaches alone might not be enough to raise the level of security sufficient for the needs of the network society. Contracting for the delivery of COTS software typically uses standard terms that are handed on take-it-or-leave-it basis. The room for individual contracting is very limited. In addition, the normal legal presumptions of approximate equality of bargaining power and comparable sophistication in evaluating benefits and risks

---

[1032] As has been argued in this study, under the principles of the private law of contract (such as rules on defectiveness) buyers' have a reason to expect certain level of security from software products. This is why courts should read in some basic security requirements into contracts that are silent on the specific information security issues.

are simply not correct in the context of software security. A typical purchaser, excluding only the largest industrial buyers that have the resources to use the expensive and scarce expertise, is only slightly more experienced than a consumer and should not thus be treated as having sufficient information for rational decision-making on costs of software vulnerabilities[1033].

> Buyers' (including industrial buyers') possibility to assess the vulnerability of the software and conduct security testing is reduced. This is not just due to the access to source code issues, but also due to the diminishing expertise of software development among IS developers (increased use of components and outsourcing instead of in-house development). All but the largest and most capable buyer organizations lack the resources or expertise to evaluate the security of a software product. Even though the higher stakes involved for industrial buyers justify measures to avoid the common information-deficiency problems that consumers face, the outright unavailability of information on the quality aspects of security together with the high costs of acquiring it due to the lack of methods to gather such information and the underdeveloped state of metrics in the quality aspect of security means that not even industrial buyers have sufficient information for sound decision-making. More importantly, no reasonable and knowledgeable person would expect them to be able to do so. Buyers that search for information only as much as it is justified on cost-benefit basis cannot be expected to use scarce and expensive experts to gather information that the COTS software vendors could much more easily deal with[1034]. In the case of private consumers the information acquisition is even more constrained.

---

[1033] Even though industrial buyers are in a better informational position to evaluate the products and vendors claims than consumers, they still are at the mercy of the vendor largely due to the lack the means (not only are white-box methods largely unavailable, but also the tools for testing even on the developer side are relatively undeveloped and not in wide use) and financial resources to evaluate the products security (high costs involved).

[1034] Exceptions are buyers with special needs for security, such as many mission-critical governmental systems. Many of the big established COTS vendors have also started to license their source code for security purposes for these customers.

Standards negotiated in individual contracts are thus unlikely to take a balanced approach to security in any time soon. At least not as long as the standard conditions for contract are drafted unilaterally by software vendors or by industrial organisations representing only them. In order for the security interest to be taken into consideration, standard conditions for COTS software contracts would have to be drafted at least bilaterally with the representatives of the buyer community[1035]. The negotiation force of the customers alone is not sufficient to shift the current balance under which COTS software vendors can get away with several security related vulnerabilities in their systems.

This is why liability rules need to be combined with the capacity to supervise these endorsed standards. Since the purchasing party does not have sufficient information to bargain for terms that express her security interests, it is doubtful that self-regulation can achieve a socially efficient standard. The self-regulative approach needs external controlling by courts in order to ease the problems caused by the informational asymmetry.

The supervisory role of courts in needed also to redress the other predictable problem of self-regulation, i.e., the failure to take into account externalities[1036]. This is especially important in the case of software security. Due to the interdependent nature of network security and the negative externalities created by security related vulnerabilities, not even the demands for more secure software made by a sufficiently big group of the largest and most capable user organisations might be enough to leverage the current practices disfavouring security in software development. Even though it can be assumed that vendors have to respond by providing more secure software if, at the margin, a sufficiently big group of the largest and most capable user organisations start to use contractual clauses setting the vendor liable for security related defects in delivered software

---

[1035] Even though this often is the case at least in the Nordic countries, the contract templates typically do not take software security issues into wider consideration.

[1036] This has been emphasised by Hugh Collins in *Regulating Contracts*, p. 302.

product, it is still questionable whether the security level would rise to be sufficient for the society as a whole. The private interest of the rationally self-interested purchasing companies might not alone, without the supervisory role played by the courts, be enough to improve the level of security in software to meet the needs of the network society.

Dispositive product liability rules, such as those in general tort and contract doctrines, include the potential to regulate secure software development and to raise the level of security in software to the socially optimal level. Improvement through changes in the reasoning in private law to being more concerned with the social and economic effects of its interventions, which have already happened to a degree due to welfare regulation, could provide a way to develop private rules applicable to enhance secure software development. The relevance of non-legal reasons, i.e., the empirical information about the impact of regulation and how the subjects of regulation have altered their conduct to adjust regulation, would have to be increased in the legal reasoning process. Following Hugh Collins, these non-legal reasons could provide the argument for fresh private law regulation by means of reinterpretation of the principles of private law[1037].

Stronger emphasis of content-based considerations in the conditions according to which the standard terms become part of the contract is one such development that holds the potential to alter the situation[1038]. In this approach towards the binding force of

---

[1037] Collins, *Regulating Contracts*, p. 52.

[1038] This is part of the development in the legal doctrine towards the protection of the weaker party. It is a reaction to the widening acceptance of the validity of standard form contracts in general. Another development in the doctrine for standard form contracts, which for its part balances the interests of the vendor and the buyer, is the interpretation of unclear standard terms to the detriment of the party who drafted them. Also the doctrine that perfectly valid liability disclaimers and exclusions clauses in standard form contracts are interpreted narrowly, i.e., in a way that limits the liability as little as possible, for its part balances the interests. Of these in the Finnish doctrine, see Hemmo, *Vahingonkorvausoikeuden oppikirja*, p. 206; Hemmo, *Sopimusoikeus II*, p. 75; Wilhelmsson, *Vakiosopimus*, p. 55. Note that especially the effects of the last two of the doctrines are easily circumvented by slightly changing the contract

standard form contract terms, comprehensive considerations taking into account the characteristic features and requirements of the type of contract and the economic activity in question are used instead of formal criteria such as the manner or form in which the terms are presented[1039]. This content-based approach is visible especially in the doctrine of surprising and severe standard terms according to which terms that significantly increase one party's obligations in comparison to default rules in indispositive law or to what is normal and ordinary for the type of contracts in question have to fulfil more stringent criteria in order to become binding[1040].

Similar effects could be induced when courts apply their possibilities to adjust and invalidate unfair, unreasonable, or unconscionable terms[1041]. Especially when the satisfaction of fairness or reasonableness is not determined solely with reference to the contractual clauses, but is determined in the entire context of the circumstances as it ought to be. This could be the case even in enterprise relations if the material justifications for the use of the terms on software business are not accepted anymore[1042]. A widely published

---

terms.

[1039] This has been presented especially by Thomas Wilhelmsson in *Vakiosopimus*, p. 88 onwards.

[1040] This is the idea of "semi-compulsoriness" of default rules as presented at least by Juha Pöyhönen, The Law of Obligations, p. 91-92, and Mika Hemmo, *Sopimusoikeus II*, p. 293, in relation to Finnish contract law and by Jan Hellner, *Lagstiftning inom förmögenhetsrätten*, p. 160, in relation to the Swedish contract law.

[1041] The application of provisions that permit judges to invalidate unfair, unreasonable, or unconscionable terms is a dominant pattern in the European regulation of standard form contracts issued to *consumers*. Main provisions can be found from the EC Directive 93/13/EEC of April 1993 on unfair terms in consumer contracts, Official Journal L 095, 21.4.1993, p. 29-34

[1042] This is so at least in countries where there are similar rules for the regulation of contracts between enterprises as there is for contracts between consumers and enterprises. This is the case, for example, in the Finnish doctrine where the Act on the Regulation of Contract Terms Between Enterprises (1062/1993) [L elinkeinonharjoittajien välisten sopimusehtojen sääntelystä] make it possible for the Market Court or even for the Consumer Ombudsman to forbid enterprises, accompanied with a threat of a fine, to use

court case establishing the abolishment of the most extensive product liability disclaimers in the case of security could alter the incentives for secure software development.

A clarification, restatement or even the promulgation of a totally new standard for conduct in a product liability lawsuit[1043] and the subsequent publicity given to it could also ease the uncertainty as to the very basis upon which a court may decide cases involving security related vulnerabilities in a software product. Already a clarification of the availability of the remedies provided by product liability rules when insecure software results in damages, could be enough to induce software vendors to re-examine their development practices by stating the obligations of software vendors[1044].

A change in the level of security provided could thus be induced by a court disallowing the use of extensive liability disclaimers and exclusion clauses on the basis of insufficient justifications[1045]. In the

---

contract terms that are considered unfair against another enterprise.

[1043] Recent argumentation in liability law also in the Nordic doctrine, especially by Thomas Wilhelmsson in *Senmodern ansvarsrätt* increasingly recognises the possibility of constructing new standards (norms) via litigation as a regulatory mechanism. Preventive effect of liability litigation is not related solely to the breach of rules explicitly set elsewhere, but litigation can also create altogether new rules (Wilhelmsson, *Senmodern ansvarsrätt*, p. 53). However, as Wilhelmsson points out in *Senmodern ansvarsrätt*, p. 186, liability litigation used as a tool for micropolitics in this new rule generating form involves numerically marginal cases since majority of tort litigations involve small damage cases with more common issues.

[1044] Influence on the behaviour of other vendors than just those represented in the litigated case would require high media attention. This could arouse public interest into the specific software security issues in a fragmented society and to feed the necessary moral discourse. This learning process in the form of moral education through litigation, depicted in the Finnish doctrine especially by Thomas Wilhelmsson in *Senmodern ansvarsrätt*, p. 103-134, can also happen without much publicity through the system of precedents or through information provision to the legislators and other regulators (Idem., p. 161-162), but at least the precedent system presumes the use of expensive advice of expert lawyers from specific jurisdictions to gather any useful guide about the requirements of the standard

[1045] Since the contractual practice in the software business categorically deviates from the default rules provided in contract and tort law, the courts could

consideration of the binding force of surprising and severe terms the factor that is of crucial importance is whether there has been a well-founded reason for the use of such terms based on the business activity in question. Several such reasons have been given and accepted for the use extensive liability disclaimers and exclusion clauses in the software business[1046]. However, the acceptability of these liability disclaiming and excluding clauses in relation to security related vulnerabilities can still be questioned.

The objective of clarifying the legal status of the parties, which is one of the objectives of standard form contracts, is naturally acceptable in the situation where the law is as equivocal as the product liability rules are in relation to software products. However, when the private regulation via contracting leads to situation that is more unfavourable than what would be the result under dispositive contract and tort law rules, the use of extensive liability disclaimers and exclusions clauses in the case of security becomes questionable. Especially because the legal status of the vendors is unilaterally improved in their favour.

Especially questionable is the argument that liability disclaimers and exclusion clauses should be allowed in the software industry due to their contribution to innovative activity in young markets. As has been argued[1047], not all innovation would be negatively affected. Only at high levels of liability costs is liability likely to reduce innovative activity. At lower levels, product liability costs provide incentives for product safety improvements and increase product R&D intensity. However, it is unlikely that product liability costs would be unbearable for the software markets. The many excuses and defences provided by product liability rules combined with the institutional and procedural hindrances for product liability lawsuits together lower the true liability risk.

---

require good justifications for the deviation from such default allocations that involve significant policy choices behind them.

[1046] Of the reasons for the almost universal use of contractual liability disclaimers, see heading 5.3.2 above.

[1047] See heading 5.3.4 above.

In addition, the argument that efficiently minimised liability risk enhances customer choice by allowing willing customers to take the risk of using possibly unfinished products and to take the possibly highly useful piece of software into use earlier becomes questionable when the quality aspects of security are considered. As has been argued several times before, even those customers that would be willing to take this risk, cannot separate between insecure and secure products. This means that their risk taking is not based on sufficient information, and they may take a level of risk that they were not willing to take, or pay a price that does not reflect the risk associated to the software in question. Even more importantly, they may consume an insecure product in such high amounts that severe damages to information security of third parties are caused.

This does not mean the clauses excluding or disclaiming liability should not be accepted at all. All that it hints is that software vendors ought to be allowed to exclude their liability for security related defects in software only to the degree that the costs would be unbearable. As Dominic Callaghan and Carol O'Sullivan argue in relation to U.K. law, clauses that limit rather than exclude liability and that are capped at an amount that is justifiable rather than arbitrary are more likely to be enforced[1048].

For example, a cap that is proportionate to the value of the contract and the amount of the supplier's insurance policy is likely to set a balance between the incentives to release new products that imply

---

[1048] Callaghan and O'Sullivan, Who Should Bear the Cost of Software Bugs?, p. 58, with reference to the case Sam Business Systems Limited v. Hedley and Company [2002] EWHC 2733. They consider the U.K. Unfair Contract Terms Act 1977 (UCTA) that includes a test akin to reasonableness in order for any contract clause attempting to exclude or restrict liability to be enforceable. Section 11 of the UCTA requires a number of factors to be considered (bargaining position of the parties, whether the goods are bespoke, the resources of the parties and the cost and availability of insurance. As Callaghan and O'Sullivan point out in Who Should Bear the Cost of Software Bugs?, p. 59, footnote 17, courts have in practice considered these factors in any situation where UCTA has required a test of reasonableness to be applied, even though the Act only requires them to be taken into consideration in respect to specified sections of the Act.

significant liability risks and the incentives to develop more secure software. In Finland, for example, clauses that limit the liability for direct damages to 15 percent of the price of the delivered software, and disclaim all liability for any indirect or consequential damages have been accepted as part of the multilaterally drafted IT2000 Terms and Conditions for IT Procurement templates. In consumer contracts, however, the limitation of all indirect or consequential damages might not be valid due to the mandatory provisions entitling consumers to compensation also for indirect losses[1049].

As the cyber-insurance industry matures, the limitation of the liability to a cap that is proportional to the value of the contract and the amount of the supplier's insurance policy could prove to be an efficacious way of encouraging secure software development without diminishing the motivation to introduce potentially highly useful new software products to the market overmuch. Due to capabilities of insurance companies to pool information, together with their superior expertise in assigning proper prices to risk and in developing safety standards, increase use of insurance could be a way to achieve the socially efficient level of care among software vendors[1050].

This might, however, be an overly optimistic view of the capabilities of the insurance companies. As has been argued above, the current level of uncertainty under traditional insurance policies combined with the unavailability and high costs of cyber-insurances results in under-investment in insurance, and thus an insufficient

---

[1049] For example, the indispositive rules in the Finnish Consumer Protection Act 1978/38 section 20, according to which the buyer is entitled to compensation for indirect losses if the defect or loss is due to negligence attributable to the seller or if, at the conclusion of the contract, the goods differed from an express representation of the seller, cannot be contracted to the detriment of the buyer. Section 20, subsection 1, of the Consumer Protection Act 1978/38 reads as follows: "The buyer shall be entitled to compensation for loss that he/she suffers because of a defect in the goods. Indirect loss, referred to above in section 10(3) and (4) shall, however, be compensated by the seller only if the defect or loss is due to negligence attributable to it or if, at the conclusion of the contract, the goods differed from an express representation of the seller."

[1050] See heading 5.3.4 above.

amount of profit-smoothing by firms and an inefficient level of risk-sharing throughout society[1051]. It is true, however, that the turn to the insurance industry can at least help in setting more specific regulatory standards and make it easier to find the socially optimal level of care. This requires a significant improvement in the availability and prices of cyber-insurance. The insurance industry would have to mature a lot in relation to the liabilities resulting from software defects.

Further U.K. development, where courts are more willing to allow suppliers greater scope for limiting liability if the supplier undertakes to fix free of charge all bugs during an acceptance/warranty period coupled with money back guarantee during that period, is not favourable in terms of incentives for secure software development[1052]. Such an approach would only mean that the current practice of report-and-patch or even penetrate-and-patch would be accepted and no incentive of original secure software development given. Even though Dominic Callaghan and Carol O'Sullivan correctly point out that the adoption of the test akin to the 'satisfactory quality" for all software, regardless of mode of delivery would offer benefits to both suppliers and customers, their normative argument in favour of a greater scope for limiting liability if an acceptance/warranty period is in place turn out to be harmful in terms of security[1053].

---

[1051] This argument is made by Jay P. Kesan, Ruperto P. Majuca and William J. Yurcik in the The Economic Case for Cyberinsurance, p. 20.

[1052] E.g., in Sam's case (Sam Business Systems Limited v. Hedley and Company [2002] EWHC 2733) the presence of such clause led the Court to conclude that it was reasonable in the circumstances of that case for the supplier to exclude all liability other than that arising from misrepresentation. Dominic Callaghan and Carol O'Sullivan in Who Should Bear the Cost of Software Bugs?, p. 59, advocate even an approach where the presence of an undertaking to fix all bugs during an acceptance/warranty period, without the money back guarantee requirement, should logically also give the supplier some scope for reducing their exposure. Whereas Sam's case involved a contract between two commercial parties, Callaghan and O'Sullivan advocate a similar approach even to the assessment of reasonableness of exclusion and limitation of liability clauses in contracts with consumers.

[1053] Callaghan and O'Sullivan, Who Should Bear the Cost of Software Bugs?, p.

> Callaghan and O'Sullivan argue that a general test that is not dependent on the mode of delivery, together with a greater scope for limiting liability if an acceptance/warranty period is in place would offer benefits to both suppliers and customers in that it would encourage suppliers to incorporate such acceptance/ warranty periods in their contracts and that would encourage them to (a) spend more time removing bugs prior to releasing software, (b) promptly rectify bugs that are uncovered after software is released, and (c) clearly explain in simple terms in the contract specifications what the software will and will not do. The fault in the influencing logic of (a) and (b) is that, since patching is such a cheap way to remove bugs, the incentives for original secure software development would be minimal. The penetrate-and-patch approach simply is negative in terms of secure software development. In the case of (c) it has to be noted that in the case of COTS software contracting the specifications are likely to be uninformative due to their standard form nature. In addition, the licenses are rarely read and understood.

This type of reasoning pattern in private in private law could enable it to develop capacity for setting standards that can truly guide behaviour. However, private law would have to forgo the flexibility of abstract principles in favour of a more explicit rationalisation of the policy behind the law. The policy statement combined with general principles could then produce more determinate guidance, which could be generalised across a series of transactions. The clarity of the regulatory standard supplied by this type of private law reasoning depends upon the explicit articulation of the policy objectives of the rule imposed by the decision.[1054]

---

59-60. Note that Callaghan and O'Sullivan concentrate on the quality aspects of software bugs and only briefly refer to the security aspects; they do not deal with the special issues of secure software development and especially not the quality aspects of security (i.e., the security related vulnerabilities).

[1054] This is what Collins in *Regulating Contracts*, p. 53-55, refers to as productive disintegration of private law where the former general rules of private law are replaced with more contextual rules informed by a more purposive approach to regulation. At the same time the private law becomes more refined in its differentiation and normative orientation. In other terms, private law in its application learns from its environment about the effects of its rules on the

Naturally, it can be doubted whether courts posses the expertise and information to device general, stable regulations that provide adequate guidance for software markets[1055]. They typically do not have information on the operation of software markets and of the possible effects of regulation. However, private law system of regulation has means to acquire the needed information. Procedural rules, for example, that permit amicus curiae for interested pressure groups and use of experts in litigation combined with encouragement for their usage are especially helpful in order for the courts to be able to set more specific regulatory standards[1056]. This would give courts access to empirical and more systematic information on the operation of markets and of the possible effects of regulation. Courts would thus be enabled to better comprehend such information in order transform it into practical regulation of such a highly complex issue as insecurity of software.

Another mechanism to improve the guiding function of private law rules is to combine liability rules with reflexive strategies towards industrial standardisation, so that it derives its quality standards from the sectoral product specifications and quality assurance processes. The courts can incorporate the quality standards set through technical specifications of the software industry as part of their standard-setting function, so that satisfactory quality and security are always defined in part as compliance with industry technical standards.

---

subjects of regulation, which then permits further refinement of the legal rules in order to modify their effects. A typical example can be found from consumer transactions where private law rules have been adjusted in a ways that incorporate the kinds of consequential considerations of policy that lead to the creation of these categories in economic and social regulation.

[1055] Hugh Collins, *Regulating Contracts*, p. 82-87, provides a deeper analysis of the expertise of ordinary courts in setting private law standards.

[1056] The use of amicus curiae might not be all that helpful in the case of software security in countries like Finland where there are no real active pressure groups that could enrol the experts of secure software development in the field. In such countries expert testimony would remain at the hands of the representatives of a specific litigation.

A variety of such technical specifications exist in the information security arena in general and in software security especial[1057]. There is no lack of standardisation efforts. Instead, the great number of competing standards and specifications has lead to fragmentation of the market and to non-interoperable solutions. As the European Commission correctly pointed out in 2001, there still is need to coordinate the current standardisation and certification activities[1058]. In addition, these technical standards are the product of market actors and, for example, might not consider externalities. A court should always be empowered to require a higher standard, or at least to insist that mere compliance with the industry standard does not guarantee that satisfactory goods have been supplied[1059].

Under this light we could expect interpretations of private law rules that take into account the specific circumstances of software business in relation to security[1060]. However, this is not likely to happen anytime

---

[1057] For a somewhat comprehensive compilation of technical information security standards from the United States in 2004, see NCSP, *Technical Standards and Common Criteria*, Appendix C. A roadmap for the ICT security standards is provided by the International Telecommunications Union's Telecommunication Standardization Sector (ITU-T) where both the standardization organisations, approved standards and standards under development and new proposed standards are presented. See the ITU-T Study Group 17 ICT Security Standards Roadmap v1.0, November 2005, at http://www.itu.int/ITU-T/studygroups/com17/ict/index.html [updated 25.1.2006, visited 1.2.2006]

[1058] Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of Regions of 6.6.2001, Network and Information Security: Proposal for a European Policy Approach, COM(2001)298 final, p. 18.

[1059] This requirement is presented, for example, by Hugh Collins in *Regulating Contracts*, p. 295-296.

[1060] There is already a class action suit pending at the State Superior Court in Los Angeles, California, USA against Microsoft for its allegedly insecure software and failure to provide adequate notice of the vulnerability of online data and of information transmitted through Microsoft operating system (Hamilton vs. Microsoft). According to the complaint, which is available at http://www.sans.org/resources/mscomplaint.pdf [7.5.2006] someone obtained unauthorised access to the social security number and bank account information of Ms. Hamilton and the proposed class representative. The data

soon. Litigation is not a likely regulatory instrument for product liability in Europe. Not only is private law reasoning still so deeply rooted in abstract principles that outright policy considerations and regulation on policy basis seems far fetched, but it also tends to favour corrective justice arguments instead of preventive and consequential claims. Private litigation is neither seen as powerful means to monitor overall product safety or security nor to send indirect behavioural requirements[1061].

As noted by Mathias Reimann, in addition to the institutional and procedural realities that seriously discourage vigorous use of product liability remedies in most European states, such as the burden of having to proof defect and/or causation even in product liability claims involving strict manufacturer liability[1062], there is also an attitudinal factor[1063]. From the European point of view, the purpose of private litigation is limited to compensation in individual cases.

---

was stored on the computer of the victim of the identity theft.

[1061] This point has been made by two leading experts of consumer law in the European Union, Geraint Howells and Thomas Wilhelmsson in 1997 in their article EC and US Approaches to Consumer Protection – Should the Gap be Bridged?, p. 263.

[1062] As Mathias Reimann, Product Liability in a Global Context, p. 150, notes in analysing product liability rules worldwide, even strict manufacturer liability is often not much of an advantage for the injured parties. The plaintiff must still show defectiveness and cause, and not having to prove fault is frequently a small benefit at best. The purchasers of COTS software are unfairly disadvantaged even by the burden of having to prove defect and/or causation in product liability claims. The concern mainly arises from perceived difficulties in proving claims due to a lack of legal or other resources needed to investigate them properly, or to an inability to gain access to essential information in relation such a complex technical product as modern software products that often involve complicated injuries.
This pretty much downplays the capacity of private law, emphasised by Hugh Collins in *Regulating Contracts*, p. 93, in relation to the private law of contract as a system of regulation, to overcome or diminish many of its structural weaknesses by using the burden of proof for the purpose of detecting violations for the regulatory standards.

[1063] Reimann, Product Liability in a Global Context, p. 152-153.

It has little, if anything, to do with larger safety or security issues[1064]. If product liability litigation has been scarce in the EU, then its use as a regulatory tool has been even more so.

The problems of private law enforcement are especially problematic because neither do the non-legal sanctions such as refusal to do business and damage to reputation and associated costs that are not directly linked to the legal liability, on which regulatory compliance typically rely on more than on the enforcement of legal rules, seem to be able to provide sufficient incentives for secure software development. The non-legal sanctions and the costs associated with insecure software have not been sufficient to improve the level of security in software.

As has been pointed above, the legal process or any other institution that is respected by the trading community as an impartial and reliable finder of fact or truth (e.g., a consumer ombudsman, a committee of a trading association or an arbitrator) does, however, contribute to the imposition of informal sanction. They function as an organised reputation mechanism collecting and disseminating what participants in the market regard as reliable information about trustworthiness.[1065]

This informative function of the authoritative declarations about trustworthiness such as in the case of a court determining that a person has broken a contract could serve as a catalyst for change in the software business. However, in relation to the contracting in the software business such a catalyst could come, instead of a court, from

---

[1064] As Hugh Collins points out in *Regulating Contracts*, p. 292, even though private law as a system of regulation has the capacity to introduce externalities into its standard setting, there plainly is a danger that in the enforcement of standards private parties will have no incentive to monitor compliance and to compel conformity.

[1065] As Hugh Collins points out in *Regulating Contracts*, p. 123-125, this type of contribution can happen even without ever invoking the power to impose legal sanctions. The influence upon contractual behaviour stems instead from the independent reactions of participants in the market to the authoritative judgements about business reputation issued by these respected bodies. According to Collins this informative influence is more important in liability law than legal sanctions.

an ombudsman or a consumer complaints board or any other organ that is more easily reachable and quicker to respond than on ordinary court. Arbitrators in the current form do not seem to fulfil this requirement. Their decisions are not public and the process involves huge costs.

Even if consumer ombudsmen and consumer complaint boards could be better institutions in the future to enhance the use of non-legal sanctions, their use in a business area that is essentially international and the issues to be solved highly speculative could turn out to be less efficacious. It can be doubted whether they are able to justify their decision and equipped to publish their decisions in a manner that would have real impact on the international software business. The national varieties in the availability of such organs further hinder their informative role. At least, much deeper international cooperation would be required in the area of software contracting

*Possible legislative actions?* If were private litigation an unlikely means for raising the level of security in software to be sufficient for the needs of the network society, then what about legislative action? A legislative action changing substantive law in a way that would restate the rights of users for compensation in software product liability cases could also alter the contractual practice towards a more balanced approach to vulnerability issues similar to a widely published court case.

Already a legislative amendment stating at least that the lack of safety test provided in product liability rules or a test akin to the satisfactory quality/fitness for purpose test in contract law ought to be the criterion for defectiveness that would be applied to COTS software as a whole, regardless of the mode of delivery, would bring clarity to the standard of conduct applied to secure software development. A clear statement of the applicability of product liability rules would significantly diminish the uncertainty surrounding the issue, and at least lower the psychological threshold for litigation for its part and thus improve the efficacy of product liability rules.

As has been argued, the standards of satisfactory quality/fitness for purpose or use and the lack of safety test provided still do not bring much specificity to the expected behaviour. However, more specificity to the standard of conduct in software development could also be provided via legislative action.

Although a more precise standard is a tempting solution for the problems in the directive force of private law system of regulation, the US experience with Uniform Computer Information Transaction Act (UCITA) suggests that it may not result in any greater clarity[1066]. As such a more precise legislative standard for conduct in secure software development thus might not provide added value to the efficacy of product liability rules. Especially because the current tests for lack of safety in the European Product Liability Directive (85/374/EEC)[1067], followed by most jurisdictions with a special product liability regime, and for satisfactory quality/fitness for purpose or tests akin to them in general contract law, have the advantage of having been considered in case law in many countries[1068].

If a legislative action is taken, the best solution may be for a statute to apply the *lack of safety test* provided in consumer product liability law for all software regardless of the mode of delivery rather than

---

[1066] Even though UCITA was drafted by the U.S. National Conference of Commissioners on Uniform State Laws (NCCUSL) with the need for clarity in the law in mind, as the Prefatory Note and many comments in final version of UCITA states (The Act that is available at http://www.law.upenn.edu/bll/ulc/ucita/2002final.htm [2.1.2006]), it is not likely to result in much more clarity as Cem Kaner, among others, argues in *Software Engineering and UCITA*, Chapter VI "UCITA interferes with the practices of independent software engineers and small consulting firms", section 8 "UCITA'S Drafting will Drive Up Businesses' Legal Expenses".

[1067] Council Directive 85/374/EEC of 25 July 1985 on the approximation of laws, regulations and administrative provision in the Member States concerning liability for defective products, OJ L 210, 7.8.1985, p. 29-33, amended by directive 1999/34/EC OJ L 141, 4.6.1999, p. 20-21

[1068] Even though litigation may not be enough to bring the required specificity to the standard due to the difficulties in generalising the standards produced by case law (they are so closely tied to the facts of the particular case that the advice of an expensive expert lawyer is needed), there is at least some experience of the ways to apply the standards.

attempting to introduce a more detailed specific standard against which software could be judged[1069]. The need to apply the lack of safety test instead of the test for satisfactory quality/fitness for purpose or use derives from the need to correct both the problems resulting form the asymmetry of vulnerability information and the lack of consideration for externalities. Since contracting seems to be inefficient in allocating the risks caused by security related vulnerabilities mainly due to the several informational problems and the inability to consider externalities[1070], there is a need to apply an objective test instead of a criterion that relies heavily on the terms of the contract between parties. This is exactly the additional value that the lack of safety test provides when compared to the satisfactory quality/fitness for purpose or use test[1071].

---

[1069] Note that secure software markets might (and are likely to) be in the state of disequilibrium due to the increasing awareness and increasing customer demand. In this situation extreme care has to be taken not to set standards on levels that might prove to be harmful under the new equilibrium.

[1070] See sections 2.7 and 2.8 above. As Hugh Collins points out in *Regulating Contracts*, p. 70, it is especially the weakness of the private law contract system of regulation in considering any third party spill-over effects that often provides the justification for much more interventionist styles of regulation that are designed to compel the parties to a transaction to internalise the costs of spill-over effects such as risks to the safety of other persons or the security of their property. Even though Collins advocates the capacity of private law of contract to incorporate considerations of externalities in its formulation of regulatory standards and the application of its remedies, e.g., by regarding the parties to the transaction as representative members of a group such as consumers, in the case of contracting for secure software litigation is unlikely to be effective against the use oppressive license terms in software contracts. Their use continues even after courts have considered them as void due to their post-breach bargaining power function.

[1071] Whereas the criterion of fitness for purpose and other similar constructions rely heavily on the terms of the contract between parties, the criterion for defectiveness in the European Product Liability Directive ("lack of safety") and similar construction in jurisdictions that follow its example emphasises the importance of the safety expectations of the consumers in relation to the possible personal injuries or damages to private property resulting from defective products.

The use of indispositive rules is indispensable for the protection of the weaker party in software contracting. As has been argued above, the unequal bargaining power combined with the asymmetrical information positions puts the buyer of COTS software in so much a weaker position that the use of default negotiable rules is not sufficient[1072]. Dispositive rules that can be circumvented via contracting are not enough to alter the current practices in software development towards more security consciousness.

> The same normative argument in favour of the satisfactory quality/fitness for purpose test has been made also by Dominic Callaghan and Carol O'Sullivan[1073]. Since they concentrate on the quality aspects and only briefly refer to the security aspects of software bugs, they miss the special issues of secure software development and especially the quality aspects of security. They do not acknowledge that contracting seems to be inefficient in allocating the risks caused by security related vulnerabilities mainly due to the several informational problems and the inability to consider externalities.

This would provide additional incentives for secure software development. Unambiguous statements about the impossibility to limit liability by contractual means under mandatory product liability rules, like EC Product Liability Directive (85/374/EEC) and those in national sale of consumer goods laws[1074], basically provide

---

[1072] The normative argument favouring the use of indispositive rules in the regulation of information security in consumer markets from the Nordic point of view has already been made in 1997 by Ahti Saarenpää and Tuomas Pöysti in their research report for the Finnish Ministry of Finance on the regulation of information security in Finland (Saarenpää and Pöysti, *Tietoturvallisuus ja laki*, p. 526-527).

[1073] Callaghan and Carol O'Sullivan, Who Should Bear the Cost of Software Bugs?, p. 58-59.

[1074] According to the article 12 of the EC Product Liability Directive (85/374/EEC) "[t]he liability of the producer arising from this Directive may not, in relation to the injured person, be limited or excluded by a provision limiting his liability or exempting him from liability." Also the Finnish Consumer Protection Act (38/1978) chapter 5 on sale of consumer goods, as

consumers with the knowledge about their rights to compensation and equips them with a greater potential to shield from the psychological effect of the liability disclaimers and exclusions clauses in software licenses[1075].

Unfortunately, neither do the indispositive laws of many countries that focus on protecting the consumer in the cases of defective products and unfair contract terms seem to be able to alter the diluted incentives. This is due to the strong post-breach bargaining position provided by the liability disclaimers and exclusion clauses strengthened by the license instead of sale argument present in typical COTS software licenses[1076]. As the post-breach bargaining power argument emphasised by Hugh Collins shows, such regulation might not be

---

an example, states in section 2 that "[a] contract term differing from the provisions of this chapter to the detriment of the buyer shall be void unless otherwise provided below." It is important to note the restriction on the scope of these rules. They primarily apply only to the consumer who either has bought the product or suffers damages from its use. If the software is delivered to the consumer through a channel of operators (importer, retailer etc.), the mandatory rules do not make the liability disclaimers and exclusion clauses void between the delivery channel operators. Also other product liability than that basing on such mandatory rules can be agreed upon differently, as already pointed out above.

[1075] Mandatory rules on defective and insecure products hold a stronger expressive influence than the rules that merely allow courts in individual disputes to adjust the terms.

[1076] As has been argued, contracting in the software business involve means to circumvent even the indispositive rules. When the law imposes mandatory rules that cannot be excluded by agreement (such as in consumer protection laws in Nordic countries and especially product liability regulation deriving from the EC Product Liability Directive (85/374/EEC), which command instead of serve the contracting practice), there still is room left to alter the legal status of the social behaviour in ways that avoid the application of these rules. In the software business such an effort is made by stating that software is licensed for use, i.e., a limited right to use is given, not sold and that no sale of goods occurs and no contracts for sale is drafted. If this argument is accepted in courts, which is largely an unresolved issue in many jurisdiction, the default liability rules both in general and consumer contract laws and in specific product liability rules following the EC Product Liability Directive (85/374/EEC) are inapplicable since they mainly apply to the sale of goods and require the transfer of ownership.

enough to change the current practices in software development. The use of extensive liability disclaimers and exclusion clauses would continue despite of them.

Even if the wider application of the lack of safety test provided in consumer product liability rules in Europe for all software, regardless of the mode of delivery, might not directly mean that the rules would be efficacious in inciting secure software development, at least it would cure the peculiar situation where different standards are applied depending on the mode of delivery. This would already be a significant improvement to the current situation and ease the uncertainty surrounding the standards applicable to COTS software. For its part, it would lower the threshold for litigation and also provide expressive influence. At least the threat of liability for insecure software would become little more probable for the vendor.

The argument that product liability rules ought to apply to all COTS software regardless of the mode of delivery is contrary to the current interpretations of substantive law in many countries[1077].

---

[1077] As pointed out in 2005 by two practitioners of UK law, Dominic Callaghan and Carol O'Sullivan in Who Should Bear the Cost of Software Bugs?, p. 56, this peculiar situation where the mode of delivery of software can impact the standard by which the software will be judged (e.g., whether it is some type of a satisfactory quality/fitness for purpose test in contractual liability law for tangible items or some type of a statutory requirement on the supplier to use reasonable skill and care as in the contractual liability for the supply of services) and also whether it is possible to exclude liability for failing to meet that standard (typically not in consumer contracts for tangible items, more easily allowed in contracts concerning the supply of service) exists in the private law of a number of European countries (e.g., Belgium, Denmark, Italy, Spain and the UK) more generally than just in product liability law. For example, in Finland two leading experts in product liability Thomas Wilhelmsson and Matti Rudanko still argue in 2004, *Tuotevastuu*, p. 79-81, that such intellectual product as software can be products under the definition of the Finnish Product Liability Act (694/1990) only if their physical existence is connected some tangible object such as a computer, a cd- or a dvd-rom, etc., in which case other software than operating systems that clearly are integral parts of computers, can be products under the legal definition only when they are delivered in a cd- or a dvd-rom. Note that this is a special problem in the coverage of the Product Liability Act in Finland. Software more clearly is a part of the definition of a product under the Finnish Sale of Goods Act (355/1987)

However, the provision of COTS software ought to be considered as a sale of goods in all situations and thus covered by product liability rules in general contract and specific product liability rules. Otherwise, the artificial separation of COTS software into products and services would lead to untenable situations. For example, the liability of a vendor of COTS software would differ in cases where she delivered the software through the Internet from those where a physical delivery channel with channel operators such as retailers was used.

Due to the extensive routine-like disclaiming of dispositive contract law and other liability rules, it is evident that they do not fulfill the facilitative role prescribed to them. In the U.S. this has spoken for altering the law to conform to the trade practices, i.e., to develop new default rules for transactions in computer information[1078]. However, this has not been very successful as the experience with the Uniform Computer Information Transactions Act (UCITA) testifies. It has been severely opposed to in the U.S. even after amendments in 2000 and 2002, which were made due to the criticism. So far it has been adopted only by two states (Maryland and Virginia).

In Europe, the approach is likely to be that it is the contracting practice that is flawed, instead the law. Long tradition for example in the Nordic countries in consumer protections laws is not likely to change. This is why legislative amendments applying new liability rules on software have not been presented in Europe and are not very likely in the future. The potential for change in secure software development lies essentially in the improvement of the enforcement mechanisms of existing product liability rules and in the restatement and clarification of existing rules[1079].

---

as Marko Mononen also points out in *Yritysten välinen tuotevastuu*, p. 104-105.

[1078] See the Uniform Computer Information Transactions Act (UCITA) as promulgated by the U.S. National Conference of Commissioners on Uniform State Laws (NCCUSL) in 1999 and amended in 2000 and 2002. The act and up-to-date information about its adoption can be found from the NCCUSL web pages at http://www.nccusl.org/Update/ (see Final Acts & Legislation – Computer Information Transactions Act) [17.3.2006].

[1079] An example of this approach is the Government Proposal 231/2005 amending the Finnish Communications Market Act (393/2003) where the

Even if legislative action would be considered, it would have to overcome hindrance for legislative efforts to alter the practice of extensively disclaiming and excluding liabilities stemming from the legislators desire to harmonise just the substantive law on the books instead of unifying the law in action[1080]. When the ultimate goal of legal harmonisation does not lie in the approximation of practical results, enacting uniform substantive black-letter rules is easily considered to be enough. This is understandable for political reasons and as a starting point for a longer journey, but this does not ensure that the harmonized rules have real meaning for real parties in real cases[1081]. According to Mathias Reimann such an effort would at minimum "require an inclusion of the rules' larger context, e.g., their interplay with existing regimes, and of the overall environment in which they will be employed, i.e., of the incentives, mechanisms, and chances to enforce them. As far as I can see, projects of that nature have yet to be launched.[1082]"

As a short conclusion, we can say that changes in the interpretation of rules and contractual practices, together with legislative amendments, of product liability rules seem laudable from the perspective of enhancing secure software development. Unfortunately,

---

explicit purpose is to clarify the legal relationship between a consumer and an enterprise in adherence to the Finnish doctrine of consumer protection and basic contract law principles (Government Proposal 231/2005, heading 4.1). The proposal intends to clarify the role and the rights of the consumer by restating the contractual liability rules for defective services developed in consumer protection legislation and general contract law doctrine in relation to the specific context of communications services in the Communications Market Act.

[1080] This is raised by Mathias Reimann in Product Liability in a Global Context, p. 153-154, in relation to the harmonisation via the EC Product Liability Directive (85/374/EEC).

[1081] It is no surprise then that the patchy empirical evidence agrees only on the issue that the practical effect of product liability rules have been very limited at best. See footnote 45 and surrounding text in chapter 5 above.

[1082] Reimann, Product Liability in a Global Context, p. 153-154.

they seem unlikely. Much depends on the development of cyber-insurance markets, which is slowly catching on[1083].

While concluding that the possibilities for controlling corporate conduct through liability law in practice should not be overrated, we now turn to the informational approaches and make some conclusions about their capacity to influence secure software development[1084].

## 6.2 Improving the influencing capacity of vulnerability reporting

Vendors have the primary duty to disclose the vulnerabilities existing in their software products. In most jurisdictions they also have certain limited obligations to disclose the vulnerabilities in their software products under the law. However, for long they ignored the issue and were even unresponsive even to the vulnerability reports made by the interest group.

As become visible in the analysis of the duties to disclose safety information under the law, they are by no means efficacious in relation to the security vulnerabilities in software product[1085]. The same reasons that hindered the efficacy of product liability rules in general are at play: most importantly the fact that vendors can circumvent their duties to disclose by disclaiming all liabilities in software licenses or by disguising the provision of software product as a license for use instead of selling a good. Thus, even though vendors might, and in

---

[1083] Even though Gordon et al. in *2005 CSI/FBI Computer Crime and Security Survey*, p. 2, explicate the low use of cybersecurity insurance in the U.S. as one of their key findings, they still are optimistic that it will gain momentum (idem., p. 11).

[1084] This is a generalisation of the conclusion made in relation to the capacities of tort law especially in the area of environmental protection in 1997 by Gert Brüggemeier in The Control of Corporate Conduct and Reduction of Uncertainty by Tort Law, p. 74, where informational regulatory alternatives are raised as a more promising approach than liability law.

[1085] See heading 5.3.2 above.

most jurisdiction do, have certain limited obligations to disclose the vulnerabilities in their software products under the law, the probability of the sanctions and liabilities occurring especially in the global context is so low that they currently do not have a sufficient weight to influence the incentives of vendors to disclose security related vulnerabilities.

Even though liability rules together with statutory obligations to inform are frequently called into action in favour of secure software development, the analysis made in this study shows that currently their capacity to incite information provision on security related vulnerabilities is seriously downsized. Serious amendments and different emphasis in interpretations would be required. They are unlikely in any short amount of time. A stronger influence for the legal duties to disclose is possible in the long-run, if vendors alter their contractual practices and start to assume at least some liability as a competitive factor.

Anyhow, the legal duties to disclose might be more of academic interest than of practical relevance. Majority at least of the big software houses have already started to disclose vulnerabilities. There is no clear need to strengthen the legal duties to disclose. Vulnerability reporting has already made most of the vendors responsive to reports and eager to disclose at least the vulnerabilities reported to them by the external parties. Vulnerability reporting has for its part elicited a change of perspective among the vendors. They have come to realise the potential of vulnerability reporting in helping testing and saving costs (as a valuable asset and a way to reduce costs of testing), even to the degree of over-reliance.

Even though a lot seems to expected from the product liability for defects in software, my analysis shows that it is not a likely to be an efficacious regulatory instrument. While the liability rules evolve and before they really start to affect the creation of secure information systems, if they ever do, external reporting of vulnerabilities seems to provide a feasible means to improve the current state of software security. Vulnerability reporting has been seen one of the most feasible and likely means to improve the current state of software security.

As an informational regulatory instrument it is both politically attractive, at least more attractive that many of the more intrusive instruments, and seems to involve a great influencing potential. It is also in line with the general policy approach to the regulation of information security where reliance is mainly on market forces, i.e., voluntary and cooperative action by the private sector[1086].

The attractiveness of the regulatory instrument is increased by the mutual benefits that it could provide for all of the stakeholders in the vulnerability reporting process. In theory, the interests of the regulators (reporters), the implementers (especially vendors) and the final objects of regulation (users) can be seen to be consistent. At least in the current form of consensus based responsible vulnerability reporting framework. The actions desired by vulnerability reporters (protective measures by the users and issuance of patches by the vendors) are also in the private interest of both the vendors and the users[1087].

This coinciding of interest speaks in favour of the efficacy of the use of vulnerability disclosure as an informative regulatory instrument to solve the problems of secure software development[1088]. Necessary

---

[1086] Even though this is most clearly visible in the cyber-security policy of the United States as critically discussed, e.g., by James Andrew Lewis in Aux Armes, Citoyens, p. 821-830, together with the reason for such a high reliance on the private sector initiatives in cyber-security, as a general approach the reliance on the private sector is also the starting point in Europe. However, the European approach involves a stronger role for the government as already the Commission Communication on Network and Information Security: Proposal for a European Policy Approach COM(2001)298 final testifies in emphasising certain market failures.

[1087] Summing up. If the action desired by the reporters comes through it would lead to net benefits for the society as a whole, which seem to be part of the objectives of the reporters. The reporters would also, in addition to personal gratifications, like to see the enhancement of secure software development and have better software to use. The vendors would, at the same time, achieve low-cost assistance in their software development efforts. Finally, the users would get, in addition to greater security, also costs savings in the form of higher quality software and less damages from attacks and accidents.

[1088] The theory of coinciding interest, presented in the literature on policy strategies, instruments, and styles by Evert Vedung and Frans C.J. van der

voluntary compliance is a reasonable expectation when the interests of all the parties are closely aligned. Vulnerability information is especially useful because the users wish to engage in the desired behaviour but have been prevented from doing so by ignorance or lack of pertinent information and also the vendors that produce both information and patches are also to gain. The arguments in favour of the cooperative form of implementation of the vulnerability reporting parallel with this.

However, the above analysis showed several hindrances to the efficacy of this type strategy of regulation by information. Not only do the interests vary in practice or are even contradictory in certain aspects, which seriously reduces the efficacy of non-coercive regulation by information[1089], but there are also several issues that need further research and interest from the practitioners. One of the main issues is the lack of capability to influence initial secure software development.

As has been pointed out above, the concentration in vulnerability reporting is mainly on the disclosure of vulnerabilities found after the release of a software package to the public. It affects secure development only in terms of patching (i.e., post the release of the software). Initial secure software development is not the primary concern. However, if the vulnerability disclosure is truly desired to influence secure software development this lacking needs to be redressed.

The ongoing deeper unification and codification of the different approaches to vulnerability reporting into the common policies and

---

Doelen in The Sermon, p. 107-109, partly explains when informative regulatory instruments should be used.

[1089] This is what the recent history of vulnerability disclosure testify. The reporters turned to full and immediate disclosure when they felt that the vendors are not taking their non-coercive reports seriously. Coerciveness still remains as a central part also in the more responsible and cooperative forms of reporting that have largely replaced full and immediate disclosure after the vendors' interest has increased and the process of reporting has matured. Note that empirical evidence on the influence mechanisms together with their efficacy is still largely lacking.

guidelines that are accepted by most of the stakeholders, together with the increasing reliance on them while handling vulnerability disclosure and other improvements in the vulnerability disclosure process is not enough as such. As long as the concentration is on post-release problems in the form of patching and the avoidance of vulnerabilities in the development phase is left relatively unaddressed, not even the consensus seeking guidelines provide relief to the central limitation of vulnerability reporting.

Whether or not it is possible for vulnerability reporting to really influence secure software development depends heavily on the ways the users of software products and their vendors utilise the provided information. At least three ways in which vulnerability reporting can influence initial secure software development was identified in the preliminary analysis.

*In the first*, historical data about the vulnerability of a specific software package and about the responses its vendor has given to the reporters can be useful in software acquisition. At least it can be useful in considering the switch to competing products. To the degree the responsiveness of the vendor to the reports become public, as they do when the disclosure is made directly to the public if the vendor is not responding, information about the trustworthiness of the vendor is also disseminated. Even though this data does not fully reflect the security of next release of the software package in question, it can be used as indication of the seller's attitude towards secure software development in general. In the COTS software markets this can be an influential mechanism since there are not a lot of other sources of trustworthiness of the vendor and the security of its software beside business reputation.

However, this information is not readily accessible to the (potential) buyers of software products. It is scattered in a variety of sources and is not necessarily presented in a form comprehensible to all concerned. Only large enterprises and true experts can be assumed to be able to use the general existing scattered vulnerability knowledge. However, they can be such an influential group of users that if they, at the

margin, engage in comparison shopping on the basis of the vulnerability of software products and the responsiveness of vendors, suppliers will have to respond. Not all even have to understand or reach the information, because at the margin it is enough that a sufficient group of influential users do.

Since not even this is certain, there is a need to gather the scattered vulnerability information into a more readily accessible form and by an actor that is seen as a reliable source of information on the trustworthiness of the products and their vendors. Fine-tuning this information provision with the insights from marketing research or research in the economics and accounting areas where financial disclosure has been investigated for a long time could prove to be useful[1090].

One development towards better efficacy of vulnerability reporting could be to rely more on governmental CSIRT teams or other organisations that are deemed reliable as intermediaries that collect and disseminate what participants in the market regard as reliable information about trustworthiness of software product and their vendors. Since they gather information about vulnerabilities on single sources that are relatively easy to access and are more likely be seen as impartial and reliable finders of fact or truth or as authoritative parties to judge the issue than external reporters, their possibilities to disseminate information about the trustworthiness of the software product and to facilitate the application of non-legal sanctions, like refusing to purchase, is enhanced. Especially when noting that it is not just about credibility in the eyes of the vendors that is at stake, i.e., that the vulnerability reports are heard and taken seriously by the vendors when they get the vulnerability report from a reliable source, but also the credibility in the eyes of the potential purchasers.

Better image of impartiality and reliability as finders of security related vulnerabilities is especially needed in relation to the vendors since the vulnerability reporting process involves a lot of conflicting

---

[1090] The differences in the disclosure of financial information and disclosure of vulnerability information need to be acknowledged as Cavusoglu et al. point out in Emerging Issues in Responsible Vulnerability Disclosure, p. 7.

interests and the reporters were for long seen to be acting against the vendors. More importantly, reliability in the eyes of the potential purchasers and current users is needed even more[1091]. If participants in the market do not respect the body as an impartial and reliable finder of fact or truth that can give authoritative judgements about the trustworthiness of the software product or about the business reputation of the vendors, then they are not likely to use the non-legal sanctions such as to refuse to purchase the vulnerable product. In the current high media activity surrounding software vulnerabilities, a mere hint is no longer enough to evoke an adverse reaction from customers and potential buyers.

From this perspective, the current development towards increased use of national CSIRT teams such as CERT-FI group in Finland and better cooperation at the supra-national or even international level which is one of the objectives of the European Network and Information Security Agency (ENISA) could be really helpful. They are at the position to provide the vulnerability information in a form that is likely to be perceived as impartial and reliable in the eyes of the potential vendors. Initial theoretical considerations imply that non-profit-based organisations such as the governmental CERT teams almost always perform better in terms of social welfare than profit-seeking organisations that pass the vulnerability knowledge to their clients ready pay for advance notifications and that encourage other people to submit vulnerability information to them to get paid in return[1092]. However, substantial improvement in their operation is

---

[1091] Whereas the validity of the reports is essential to the vendors, which could be improved, e.g., with a more intensive dialog between the reporters and the vendors as pointed out by Tiina Havana in *Communication in the Software Vulnerability Reporting Process*, p. 69, the perceived trustworthiness of the source of vulnerability information is more important to the potential purchasers and to the current users if to be used for comparison shopping.

[1092] According to Karthik Kannan and Rahul Telang, An Economic Analysis of Market for Software Vulnerabilities, p. 4, this is due to the possibility of information leakage in profit-based mechanism. Even if the leakage is prevented, they show that a profit-based mechanism performs better than a non-profit-based mechanism only under certain conditions (idem. p. 11). However, Hasan Cavusoglu, Huseyin Cavusoglu and Srinivasan Raghunathan

still needed. For example, even though governmental CSIRT teams and alike governmental organisations possess best capabilities to influence secure software development, they are currently unfortunately far behind non-governmental organisations in providing all the needed information quickly and correctly[1093].

Note that increased disclosure of vulnerabilities as such serves to raise awareness of security issues and can thus help in the mobilisation of social pressure even without the vulnerability information actually being used in comparison shopping. This is the intrinsic predisposition shaping effect that the vulnerability reporting can have. However, it largely depends on the degree to which vulnerability disclosure is accepted as a correct practice in information security and the way the actors perceive its role. Due to the conflicting interest and the time-to-time heading debate, special care is required from the intentional use of vulnerability disclosure as an intrinsic predisposition shaping mechanism[1094].

---

sucpect in Emerging Issues in Responsible Vulnerability Disclosure, p. 24-25, the ability of non-profit-based mechanisms to yield a higher social welfare than profit-based mechanisms because the role of the vendor is downplayed and call for future work that would consider the role of the vulnerability handling process of the vendor.

[1093] According to the analysis made by Urs E. Gattiker in CERTs, vendors, alert services – can they deliver vital information for protecting your digital assets during public holidays? EU-IST News from CyTRAP labs, dated January 4th, 2006, published in weekly online serial publication EU-IST News from CyTRAP labs (ISSN 1600-1869), 7(2): news no. 3, January 8th, 2006, available at http://security.weburb.dk/frame/newsletter/InformationSecurity/677 [10.1.2006], the ones that provided the most complete information about the Microsoft VMF (Windows Meta File) vulnerability case were, without exception, all non-governmental organisations.

[1094] For example, when the public disclosure of vulnerability is presented as an ideological argument of the freedom of information and its positive relationship to security, the intention is to shape the intrinsic predispositions of the participants in the vulnerability reporting process. However, since this ideological argument is not deemed to be correct as such, and possibly is not compatible with existing institutionalised values where some form of responsible disclosure dominates, it may thus create increased objections towards the practice of public disclosure and erode its legitimacy. The continuity and compatibility with existing institutionalised values has been

*In the second*, vulnerability reporting can influence initial secure software development more forcefully if the information provided is internalised by the vendors. This would require the vendors to distribute the information about discovered defects to their software developers and to make that information an essential part of the software development process. However, such internalisation of the vulnerability information into the secure development process is not done all that commonly[1095]. It would require a much deeper commitment to initial secure software development than is currently present in the COTS software market. Vulnerability reporting would have to initiate software houses into changing their development from release-early-and-patch-later approach to real commitment to initial secure software development.

There are, however, certain mechanisms by which vulnerability reporting could be used to initiate such deep changes in attitudes, but it is not solely in the hands of reporters of vulnerabilities. It would require a closer cooperation between the actors in the development of vulnerability disclosure policies. As was pointed out above, currently even the common policies and the guidelines for vulnerability reporting do not provide any rules, practices, procedures, or demands for the information communicated to be used for the benefit of initial

---

raised as one of the necessary conditions for the law to generate the internalisation of new attitudes implicit in the conduct required by the new law is raised by William M. Evan in Law as an Instrument of Social Change, p. 557-558. Even more important it is for the less coercive regulatory mechanisms such as the vulnerability reporting.

[1095] Only a little over half of the receivers of the reports answered in a survey conducted by Tiina Havana in 2002, as reported in Havana and Röning, Communication in the Software Vulnerability Process, p. 8, that they pass the information further to their software developers with the intention of preventing similar vulnerabilities occurring again. Of all the respondents to the survey only 15 per cent do this, and when they do it, the information does not have an essential role in the software development process. Thus, the information gathered during the reporting process is not widely utilised for the benefit of secure software development.

secure software development[1096]. By developing such requirements, the normative policies for vulnerability disclosure could be used to gain a deeper involvement in the secure software development process.

The most likely parties to develop and to be able to enforce such requirements in their policies and guidelines for vulnerability reporting are national coordinators (like US-CERT, Aus-Cert, etc.). Of the policies set up by coordinators, the objective of enhancing the public interest in information security and to aid to reach a level of security that is sufficient for the whole society is most clearly protected by the national governmental teams[1097]. National security interests, such as the protection the critical infrastructure, are also high in the list of objectives of national teams[1098]. This raises the importance of the

---

[1096] For example, there typically are no demands for the vendors to pass the information about discovered vulnerabilities to their software developers in order to prevent similar vulnerabilities in the future. See heading 4.3.3 above.

[1097] For example, the main function of CERT-FI is to promote security of the information society by assisting its clients (every finnish private individual, company or representative of the government) in preventive actions for computer security incident and the minimisation of the risk of vulnerability. CERT-FI also receives the telecommunications operators' reports on information security incidents and threats demanded by Section 21 titled ”informaton security notifications”of the Finnish Act on the Protection of Privacy in Electronic Communications (516/2004). An unofficial translation of the Act is available via Finlex at http://www.finlex.fi/fi/laki/kaannokset/2004/20040516 [31.10.2005]. See version 1.1 of the charter of the CERT-FI headings 3.1-3.4, published 7.6.2004, available in pdf -format at, http://www.ficora.fi/suomi/tietoturva/certtoiminta.htm (in finnish) [31.10.2005]. The charter is based on Interet Engineering Task Force's best practice as presented by Brownlee and Guttman, *Expectations of Computer Security Incident Response*, in RFC 2350.

[1098] See, for example, the web pages of the US-CERT, http://www.uscert.gov/aboutus.html [31.10.2005]. Such national security interest might somewhat hamper international cooperation in the area as the fear of U.S. centricity expressed in Europe show. The European Commission raised its special fear of the worldwide coordination being done through partly U.S. government funded CERT/CC and that CERTs in Europe are dependent on its information release policy in its communication on network and information

disclosure policies advocated by national coordinators, since they posses the ability to use the policy as a leverage to modify the incentives other stakeholders face for the benefit of the society at large[1099].

National governmental teams acting as coordinators and being funded by the government are also in a unique position to compel others to follow their typically more public interest biased policies. However, this is not typically done and the policies for vulnerability reporting concern only those who utilise the national coordinating teams in their reporting[1100]. National CSIRT teams and alike do not use their regulatory capacity in the public interest at the level of policies. The public interest is sought by means of informing and assistance, not by developing socially optimal reporting policies[1101].

---

security, COM(2001)298, p. 21. As a response, EU is studying ways to strengthen cooperation between EU member state CERTs and CSIRT teams and other similar organisation. The establishment of the European Network and Information Security Agency (ENISA, http://www.enisa.eu.int/) is part of this. See Regulation (EC) No 460/2004 of the European Parliament and of the Council of 10 March 2004 establishing the European Network and Information Security Agency, Official Journal L 077, 13/03/2004, p. 1–11.

[1099] Due to their role in enhancing the public interest, the national coordinating teams are in an especially good situation to develop disclosure policies that minimize total social costs, and not just private costs of a specific stakeholder. Cooperation at an international level is a minimum requirement for this to happen. The priority of governmental funded coordination under welfare economic considerations is shown by Kannan and Telang in An Economic Analysis of Market for Software Vulnerabilities.

[1100] Usually there is no obligation to invoke a coordinator in a reporting process; direct contact between the vendor, the discoverer, and the users of the product is also widely used.

[1101] Even though the national teams are in the best position to enforce socially optimal reporting policies due to their role in advancing the public interest, it is not clear that they could or even should do so. Since they act as mediators in the process, their policies would only be complied widely (assuming that the socially optimal reporting policy is not the one that gains wide acceptance by all stakeholders) if there would be an obligation to invoke a coordinator in the reporting process. Since the role of coordinators is just to assist and advance the communication and cooperation between the main actors (the vendors and

Another path would be to set requirements for the use of vulnerability information in the initial secure software development in the common policies and guidelines for vulnerability reporting. Even thought these policies and the guidelines currently provide no procedural support for this and the concentration is purely on fixing the problems that occur after release, they could be used to initiate such changes. The use of the common policies and guidelines as signals of the responsible behaviour of the vendor could be used as a leverage for the demands of deeper commitment to initial secure software development. This would, however, require substantial improvement since currently the common guidelines do not even emphasise the introduction of a vulnerability into a product and the life-cycle of a vulnerability depicted in them begins with discovery of the existence of a vulnerability in a specific software product[1102].

Also research on the ways and forms in which vulnerability info is reported to the vendors could show ways to improve the efficacy of this instrument in influencing initial secure software development. Knowledge of the parties to whom the info is reported and of the channels that are used together with the organisational position of those processing vulnerability information could show paths to improvement. For example, if the information about reported

---

the reporters), this might be desirable only if the vendors and the reporters could not cooperate in socially optimal way. More research needs to be done in order to verify this.

[1102] The formal life-cycle model of vulnerabilities presented both in the OIS *Guidelines for Security Vulnerability Reporting and Response*, p. 2-3, and in the recommendation for NIAC guidelines presented by Chambers and Thompson in *Vulnerability Disclosure Framework*, p. 13-16, begins with discovery. However, the NIAC guidelines at least acknowledges the introduction of the vulnerability into the product and then starts with the resolution of the vulnerability. Even though this could be explained by the shorter writing style of the OIS guidelines, it still represents the differences in attitudes and objectives. The major software vendors and security companies behind the OIS guidelines may not want to admit the existence of their liability for the vulnerabilities found from their software.

vulnerabilities and customer feedback, for example, about defects in functionality that more typically is made part of the software development process and especially part of the requirement of the next release, is divided and processed by altogether separate actors inside an organisation, the likelihood of vulnerability information being utilised in initial secure software development is low.

*In the third* form, the reporting of vulnerability information could also contribute to improving the engineering quality of software products, by supporting the academic and research communities' ongoing efforts to identify common security vulnerabilities, the conditions under which they occur, and methods to avoid them[1103]. The knowledge of vulnerabilities, not just of their existence but also of their details (like the specific coding or implementation error) could also be used to train current and to educate future software developers more widely to avoid the causes of the vulnerabilities. Due to the current lack of teaching material such knowledge, especially if codified, could serve as an important educational resource. For example, recent textbooks on secure software development use known vulnerabilities especially as examples to clarify and make things more concrete[1104].

Even though the influence on the initial secure software development is the most important issue to be tackled with if the vulnerability reporting is really desired to have an impact on the software development practices, there is still strong need to develop the vulnerability reporting process as such. The issue of post breach correction continues to be important for long. It is going to remain as an essential part of the general post-release improvement of existing software products. Even though external reporters of vulnerabilities at the first glance seem to provide information that the vendors ought

---

[1103] This is made explicit in the OIS *Guidelines for Security Vulnerability Reporting and Response*, p. 1.

[1104] See, e.g., Graff and van Wyk, *Secure Coding*, from 2003 and Viega and McGraw, *Building Secure Software*, from 2002.

to disclose themselves, in fact external parties are always are going to find vulnerabilities that the vendor could not have found even if it had conducted sufficient quality and security assurance. As has been acknowledged several times before, a number of vulnerabilities are always going to be contained in released software products despite the best efforts of the developers. This is why the development of the vulnerability reporting process as such is important.

However, if the lack of influence for initial secure software development is not addressed properly, vulnerability reporting is not likely to be efficacious in enhancing secure software development in other forms than just assisting in the post-release patching of software. It will not be able to alter the current high reliance on post-release correction of vulnerabilities. At the same time it may cause further problems for initial secure software development by promoting such over reliance on patching. The external discovers would continue to be used to replace the testing work that the software vendors already ought to do under product liability rules[1105]. This calls especially for more empirical research on the influence capacity of vulnerability reporting as a regulatory instrument following the research agenda for private politics presented by David P. Baron[1106].

## 6.3 The way forward

The wider regulatory approach adopted in this study showed many interesting paths that call in question the force of the law in the regulation of secure software development. At the same time it enabled the identification of the regulatory function of many issues that typically are not considered as regulation and showed that they

---

[1105] As the analysis of the default product liability rules suggested, software vendors ought to test for security and to inform their customers of vulnerabilities. These obligations are just bypassed with the current contractual practices.

[1106] Baron, Private Politics, p. 47-53.

shape secure software development much more significantly than any law currently does. This might partly be due to the youngness of the information security as an art and a science. There have not been efficacious governmental regulatory solutions in place that could have been used tackle with the problem of insecure software or, at least could have been adapted for that purpose. Vulnerability disclosure has for long been the only somewhat efficacious method of regulating secure software development in a short period of time.

At the same time it also tells about the necessary diversification of regulators and regulatory solutions especially in a rapidly developing field like the software business. Because there were no efficacious regulatory means to influence secure COTS software development, the users sought for new instruments. The original unresponsiveness of software vendors to the vulnerability reports, the easy means of communicating especially inside active groups of enthusiastic IT users, and the high media attention providing a means to inform also the general public led to an increased regulatory use of vulnerability disclosure. In the strictest sense it meant immediate public disclosure of all information.

It is questionable whether the future enhancement of such an important area of information security as secure software development could be left to such an ad hoc mechanism as vulnerability reporting. Not even with the more mature and responsible disclosure schemes. With the increasing awareness of information security issues among policy makers and legislators, and the maturing of information security as a central art and science in the future development of the networked society we can expect that legal regulation in its many forms is going to play an increasing role in the future. Already the notion of the policy evaluation studies about the escalation of regulatory instruments predicts this. So far the reliance has largely been on less coercive informative instruments.

Governmental regulation has already started to play a more important role. Not only is the avoidance of stricter and more intrusive

governmental mandates about software security a central impetus for the recently emerged more favourable approach to vulnerability reporting, but vulnerability disclosure already takes place in the shadow of government. Governments encourage such activity and support it by establishing organisations to coordinate reporting such as national CSIRT teams.

It can be expected that governmental regulation will increase in the long run. Especially if the joint governmental and private sector type of regulation does not efficaciously start to influence initial secure software development, i.e., vulnerability disclosure does not develop into the direction where it could influence initial secure software development. In other words, if vulnerability reporting as a process is not developed into the direction that it really would start to influence initial secure software development and the effects of software product liability rules remain as slight as they would seem, we can expect that stricter and more intrusive governmental regulations are enacted in the future. The regulation of secure software development is coming of age for more stricter forms of regulation.

Probably they will come in the form of sectoral product safety regulations that lay down relatively general essential requirements of security similar to the Directive 2001/95/EC of General Product Safety[1107]. The task of harmonising technical specifications and setting of more specific requirements for software security would then be left to the recognised national standard bodies and to the designated European standardisation bodies such as the CEN (European

---

[1107] Directive 2001/95/EC of the European Parliament and of the Council of 2 December 2001 on general product safety, OJ L 011, 15.1.2002, p. 4-17. Public regulation has been considered to be the best measure for product safety in Europe as noted, e.g., by Geraint Howells and Thomas Wilhelmsson in EC and US Approaches to Consumer Protection, p. 225. Private litigation in the European tradition has been considered to be limited to compensation in individual cases and has not been seen to have much to do with larger safety issues.

Committee for Standardisation[1108]), CENELEC (European Committee for Electrotechnical Standardisation[1109]) and ETSI (European Telecommunications Standards Institute)[1110]. Software developed according to these harmonised standards would then enjoy a presumption of conformity with the essential requirements.[1111]

Of course such development presumes that such standards can be developed. Research into the metrics on the basis of which software security could be measured is just emerging. Of course there are a variety of standards for information security in general as already pointed out above, but for software security they are scarce; excepting the Common Criteria which has largely been considered too expensive and bureaucratic for the evaluation of many software products[1112].

---

[1108] CEN is a system of formal process to produce standards founded by the national standards bodies in Europe. See the web pages of CEN at http://www.cenorm.org/cenorm/index.htm [13.2.2006].

[1109] CENELEC is a non-profit technical organisation set up under Belgian law and composed of the National Electrotechnical Committees of 28 European countries. See the web pages of CENELEC at http://www.cenelec.org/Cenelec/Homepage.htm [13.2.2006].

[1110] ETSI is an independent, non-profit organisation officially responsible for standardisation of information and communication technologies in Europe. Web pages at http://www.etsi.org/ [13.2.2006].

[1111] For a complete story behind the European style of product safety regulation, see the five volume series of Working Papers from the European University Institute (EUI) written by Christian Joerges, Josef Falke, Hans W. Micklitz, and Gert Brüggemeier, European Product Safety, Internal Market Policy and the New Approach to Technical Harmonisation and Standards.

[1112] Common Criteria for Information Technology Security is a standard for evaluating information technology products and systems, such as operating systems, computer networks, distributed systems, and applications. It states requirements for security functions and for assurance measures. It has been developed on the basis of several national and larger regional security evaluation criteria and there is also an equivalent ISO standard 15408. For critical analyses and suggestions for the improvement of the Common Criteria process see especially Fred B. Schneider *Trust in Cyberspace*, p. 199-209, from 1999 and the report on *Technical Standards and Common Criteria*, Appendix B, p. 2-3 and Appendix E, provided by the U.S. National Cyber Security Partnership (NCSP) in 2004.

The problems in the development and implementation of the Common Criteria and similar previous evaluation schemes also tell about the complexity and difficultness of developing standards for software security. No single technical standard suffices since the requirements differ between types of software and the context of use.

Neither can the implementation and functioning of these standards be evaluated merely with regard to their technical qualities. In order for software security standards to be a means of achieving an adequate level of security for the needs of the network society, their implementation needs to be evaluated also with regard to ethical and legal principles and in the light of the basic constitutional rights[1113].

This has, unfortunately, proved highly cumbersome. Even in Finland, where the role of information security in state action and society is relatively well developed, the review of the protection of basic rights in the regulation of information security has turned out to be a highly complex and time consuming task[1114]. This is partly

---

[1113] This has been emphasised by Henrik Kaspersen already in 1991 in one of the first conferences ever considering the role of law in securing computer networks. See Kaspersen's paper "How to Advance Computer Security by Legal Instruments?" in the proceedings, p. 92.

[1114] In the original Action Plan to implement the National Information Security Strategy, adopted by the National Information Security Advisory Board in spring 2004, there were projects targeting the safeguarding of fundamental rights as required by the Finland's National Information Security Strategy (see Ministry of Transport and Communications Finland, *Turvalliseen tietoyhteiskuntaan*,heading 4.1 at p. 31). However, already in the first progress report at the end of 2004 the Advisory Board, *Creating a Safer Information Society*, p. 9, had given priority to other projects and arranged the project aiming to ensure fundamental rights into groups with other projects, each supporting the higher priority projects. In the second progress report at the end of 2005 the Advisory Board, *Tietoturvalliseen tietoyhteiskuntaan*, p. 8 and p. 39-41, had altogether absorbed the action of ensuring fundamental rights into the larger priority project of information-secure electronic services and to the review of related legislation. More information about Finland's National Information Security Strategy and of the work of the Advisory Board, see the web pages of the Ministry of Transport and Communications Finland, at http:// www.mintc.fi/ (at the English language portal see Communications –

due to newness of the issue. What we need is general principles and deeper analysis of the role of information security among the central tenets of the constitutional state and the system of basic rights together with the place of information security among central legal principles. Only a comprehensive and systematic understanding of the versatile principles of information security and the various connections it has to constitutional rights can provide means for genuine evaluations of ensuring of fundamental rights both in the technical standards and in the regulation of information security more generally.

In line with the general argument made in theory of regulation, it is clear that no single regulatory instrument alone is enough to improve the current state of security in COTS software. The analysis has shown that software product liability and vulnerability reporting as regulatory instruments in their current form do not provide sufficient incentives to alter the practices in software development towards better security consciousness. With significant changes and in combination possibly even with other regulatory instruments they could be enough to alter the incentives for secure software development. Especially because the general awareness of security issues among the various types software users seems to be increasing.

Even though a regulatory mix where several instruments are combined probably would be the best solution, such an approach would require a coordinated policy approach. This implies a more general argument. The underlying argument of this study is that secure software development, as an important societal objective, is that kind of a conduct that would be desirable for the law and regulation in general to promote or encourage, to give some motive or ground for or, at least, not to hamper unnecessarily. As a public policy, it must be taken into consideration while drafting laws and other regulation that may have effects on secure software development and not just

---

Information Security and Privacy Protection – National Information Security Strategy) [16.3.2006].

when drafting specific standards for secure software development. If taken seriously as an important public policy, effects on secure software development would have to be raised among the issues to be considered when initiating any type of regulation that might influence software development.

This concerns especially the central property rights rules relating software. So far the effects on information security in general and on the incentives for secure software development especial have not been considered appropriately. For example, the U.S. experience with the anti-circumvention provisions in the Digital Millennium Copyright Act (DMCA)[1115] codified in section 1201 of the United States Code title 17 on Copyrights show that they have been used to stifle free speech and scientific research in information security[1116]. Also in Europe, the anti-circumvention provisions in the EU Copyright Directive (EUCD)[1117] were drafted with intense lobbying and public debate that has not ended to the EU level. Another round of extensive lobbying and resistance has appeared at the national implementation level[1118].

---

[1115] The Digital Millennium Copyright Act of 1998, Public Law No. 105-304, signed to law in 28th October 1998, implements WIPO (World Intellectual Property Organization) Copyright Treaty into U.S. Federal law.

[1116] The problems that have appeared in practice have been described especially by the Electronic Frontier Foundation (EFF) in Unintended Consequences: Five Years under the DMCA, v. 3, September 24, 2003, available at http://www.eff.org/IP/DMCA/?f=unintended_consequences. html [14.2.2006]. EFF is a non-profit civil liberties advocacy and legal organisation for the networked world based in the U.S.

[1117] Especially article 6 of the Directive 2001/29/EC of the European Parliament and the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, OJ L 167, 22.6.2001, p. 10-19. This is the EU implementation of the WIPO Copyright Treaty.

[1118] This was expected by Bernt Hugenholtz, a professor of Intellectual Property Law and director of the Institute for Information Law of the University of Amsterdam (IViR) who chairs the Intellectual Property Task Force of the Legal Advisory Board of the European Commission, already in 2000 when he wrote an opinion to the European Intellectual Property Review (EIPR), Why the Copyright Directive is Unimportant, and

The main reason for the intense outcry was the legislators' lack of consideration of consumer protection issues and concerns of basic rights and liberties especially when drafting the copyright rules. What the legislators', at least in Finland, largely ignored was the newly established importance of copyright for the functioning of the network society. Former highly specific and marginal legal issue had become a central element for the functioning of both the new e-business and the network society in general. Especially the potential chilling effects for security research in general and vulnerability research especial remained largely hidden. Also the effects of the technological protection measures to the quality and usability of informational product like software, which is a central consumer protection issue, was largely ignored. It not only concerns the right to take copies for private purposes, but also the functioning of the products altogether. Protection of the consumers' expectations of quality in relation to informational goods was largely bypassed.

Good development towards integration of the considerations for the effects on secure software development is the inclusion of the requirement to consider the possible effects of parliamentary laws on information security in general, as part of the effects on the information society, in the guidelines for the drafting of government proposals in Finland given in 2004[1119]. This holds the possibility that the effects parliamentary laws may have on secure software development are considered in the process of drafting them.

Unfortunately, the government officials drafting laws typically do not have the expertise to conduct such an analysis and, at least in Finland, the voice of the experts of software development is not present due to the lack of the use of expert advice and hearings of experts in the drafting process. Finland is also going towards the European type of lobbying in the drafting of laws and forsaking its tradition in using expert committees and councils to give advice in

---

Possibly Invalid, p. 501-502.

[1119] See Valtioneuvosto, *Hallituksen esityksen laatimisohjeet*.

the drafting of parliamentary laws. The voice of the experts in software development and information security typically does not get heard because active civil rights groups that could enrol the experts in the field are still lacking in Finland. We still do not have such influential public interest groups like the Electronic Privacy Information Center (EPIC) in the U.S. that gathers the voice of experts to draw attention on emerging civil liberties issues such as the protection of privacy. This is a crucial lacking in a situation where moral discourse in the network society and changes in the paradigms of secure software development widely remain at the shoulders of civil rights groups and alike, especially in those countries where class action suits are not available[1120].

Moreover, the requirement to consider the possible effects of parliamentary laws on information security has unfortunately remained only as a statement of the importance of information security. The drafting guidelines are not widely used in practice of law-drafting as a report published in autumn 2005 on the state and development of law-drafting in Finland shows[1121].

This calls for an emphasised role for risk analysis and management not just in software and information system development, as is currently fashionable to argue, but also in the regulatory field[1122]. At

---

[1120] For a critical overview of the Finnish consultation process see the OECD review of regulatory reform in Finland from 2003, *Government Capacity to Assure High Quality Regulation in Finland*, p. 25-29. For the European style of consultation in regulatory drafting, see the report of the Better Regulation Task Force (renamed as Better Regulation Commission in the beginning of 2006) *Get Connected*, where also the limited role of the comitology system together with the role of expert groups and advisory committees in the EU is acknowledged. At the same time the report proposes a less competitive model of consultation than lobbying where stakeholders, instead of battling for influence in order to seek favourable policy outcomes for a particular set of interest, would see consultation as a co-operative dialogue that helps to shape EU legislation in ways that maximise the benefits for all.

[1121.] Lainvalmistelun kansliapäällikköryhmä, *Tehokkaampaa, suunnitelmallisempaa ja hallitumpaa lainvalmistelua*, p. 211-212.

[1122] The central role of risk analysis and the usefulness of its methods in legal

the same time it further emphasises the need to develop comprehensive and systematic understanding of the versatile principles of information security and the various connections it has to the basic tenets of a constitutional state and the protected basic rights. Only when armed with such an understanding can genuine considerations of the effects of parliamentary laws and other regulation on information security in general be made.

---

decision-making has been recently illustrated by Peter Wahlgren in *Juridisk riskanalys*.

# Epilogue: value protection and decentred regulation

I have been arguing that such partially self-enforced regulatory instruments like software product liability and even such joint regulation like vulnerability reporting could be used to, with modifications, to achieve or at least to come closer to the security of COTS software as a social objective. In the analysis of regulatory capacities of these instruments I have abstracted away from real situations and have discussed the capacity of specific regulatory instruments without reference to the actual values[1123]. I have discussed regulation as a set of means or capacities which may be put to a variety of uses; the social objective of achieving secure software could be one. With this, I have left aside the larger question of values.

In doing so, I have rendered myself guilty of the basic vice of the discussions of the use of indirect regulatory instruments; the disconnection from the value base. When concentrating heavily on how to best design and implement policy, the normative concerns of what that policy should be or what values ought to be pursued easily become a secondary question. For example, the basic textbooks on policy analysis portray the analysis and selection of policy instruments as a rational, linear and technical exercise[1124] and in doing so oversimplifies the process of instrument choice by ignoring the important political factors involved, including ideologies and values[1125].

---

[1123] This is similar to what Michel Foucault does in his analysis of the technologies of power. This point about Foucault's work has been made, e.g., by David Garland, Punishment and Modern Society, p. 169, in a brilliant discussion of the major social theories of punishment.

[1124] See e.g., Weimer and Vining, *Policy Analysis*.

[1125] This point is made by Peters in The Politics of Tool Choice, p. 552. This dismissal of the value base is a general problem in the dominating regulatory

Value decisions, their makers and ways of doing them are, however, not only present in making the policy for regulating, but also in the choice of instruments. Because the instruments as such include certain values, to decide what instrument to use already is a value decision. In addition to the relatively clear case of influence of values to the choice of prescriptive rules and market-harnessing instruments[1126], value-decisions are also present in the choice of other regulatory tools. Cultural values are embedded in social norms. As Christine Horne points out from the perspective of sociology, one way of understanding the internalisation of norms is by individuals coming to value the behaviour specified by a norm for its own sake[1127]. When individuals internalise social norms or groups try to induce certain norms (e.g., with the use of law) they, therefore, make a choice among different values.

Values can also be embedded into technology[1128]. As the famous argument of Lawrence Lessig in *Code and Other Laws of Cyberspace* spells out (freely interpreted), the code (architecture of cyberspace) embeds fundamental, sometimes even 'constitutional', values, and to choose among variety of available code is, therefore, to make important choices between different values[1129].

---

literature deriving from the effectiveness-based critique of regulation according to which "…regulation is not achieving what it set out to achieve … usually accompanied by prescription of what techniques should be used instead, with greatest attention given in most recent literature to the 'decentred' techniques". Black, Critical Reflections on Regulation, p. 22. Similar notion is made by Tala in *Lakien vaikutukset*, p. 139.

[1126] Peters in The Politics of Tool Choice, p. 552-564, considers the effects of political decision-making on values to the choice of traditional tools like prescriptive rules and market-based instruments.

[1127] Horne, Sociological Perspectives on the Emergence of Norms, p. 4.

[1128] This argument has been made, e.g., Helen Nissenbaum in How Computer Systems Embody Values, p. 118-120.

[1129] This is a forceful argument in relation to information security. Not only are deep political controversies (concerning issues like privacy, anonymity, freedom of speech etc.) involved in the implementation of different technical or procedural ways to secure systems, but these techniques and procedures can also be used for so many purposes (e.g., to enhance national security, to enhance privacy, to protect IPRs or other corporate property).

The focus on instruments diverts attention form the issue of how the desired ends and values should be defined and by whom, despite the radical rethinking of the ways in which societal ends can be achieved that the focus on the techniques has elicited[1130]. The instrumental rationality behind most regulation tends to divert the regulatory scholars and practitioners attention away from the considerations of consistent values or of coherence of principles[1131]. Only a lip service to the need to address the substantive value concerns and to make the important political considerations when deciding on the instruments used is often made without further consideration on how to do this[1132].

The definition of regulation as an intentional, systematic attempt at problem-solving, adopted in this study leaves the door open for the discussion of the value base regulation should be subjected to and according to which it should be made legitimate and/or accountable in some way as pointed out by Julia Black[1133]. However, I am not going to go through that door in this study. The value base

---

[1130] Black, Proceduralizing Regulation: Part I, p. 598; Tala, *Lakien vaikutukset*, p. 80-81 and 139-140. This problem of instrumentalism actually is part of a more general form of critique according to which regulation is not directed at the appropriate goals, and/or that it is not being pursued in accordance with, or made subject to, certain values. Black, Critical Reflections on Regulation, p. 22. In relation legislation, see Tala, *Lakien vaikutukset*, p. 75-82.

[1131] Smith, What is Regulation?, p. 39; Tala, *Lakien vaikutukset*, p. 79. While the ideal of legal reasoning is internal coherence and autonomy from operational or outcome considerations, with regulation, by contrast, problems in widely diverse domains tend to entail quite context specific goals and techniques. The policy maker or regulationists is interested only in which rules (or other techniques) work best.

[1132] Only the role of cost-benefit analysis, regulatory impact analysis (e.g., the examination of the effects of a law reform), and the research and expertise concerning them is discussed. The message typically is limited to explicating that they cannot be a substitute for political decision-making, but rather a valuable complement to the latter. This is typical in the tools approach to governance. See, e.g., Tala, *Lakien vaikutukset*, p. 214; Peters, The Politics of Tool Choice, p. 563; Salamon, The Tools Approach and the New Governance, p. 606-607.

[1133] Black, Decentring Regulation, p. 143.

depends too much on the national context and the special circumstances of the employment of the instruments that generalisations are difficult to make. I simply confine to note that due to the possibility of avoiding political responsibility by hiding the agenda with the use of invisible instruments using indirect means of influence[1134], it is important to keep in mind that transparency is a strong argument when choosing between different instruments. It might not be the most forceful, but transparency of the instrument should weight a lot when there are alternatives to choose from.

It is, however, important to note that information security can be seen both as a collective good, a guarantee of the functioning of the network society[1135], and as a part of our fundamental right to security at the individual level; as an individual right instead of a collective good[1136]. Due to the strong relationship with constitutional rights, there is a need to elaborate a little further the question of the way values are protected and of who can protect them.

Under the strong legalistic Finnish regulatory culture it is by no means clear that other than parliamentary laws can be used when guaranteeing the use of basic informational rights. This odyssey is

---

[1134] In a constitutional democracy the regulation should be public. By using non-transparent instruments the agenda (values and ends) behind the regulation can be hidden from public scrutiny.

[1135] Of this discussion in Finland see, e.g., the most recent contribution of Tuomas Pöysti in ICT and Legal Principles, p. 560-600.

[1136] At this individual level information security protects our identity and our right to informational self-determination. Note that this is not an established interpretation. Of this discussion in Finland, see the report of the Institute for Law and Informatics to the Ministry of Transport and Communications Finland, Saarenpää et al., *Sähköinen viestintä, tietoturvallisuus ja perusoikeudet*. The report is part of coordination of the actions implementing Finland's National Information Security Strategy that has been the task of the National Information Security Advisory Board since it began its work in spring 2004. More information about Finland's National Information Security Strategy and of the work of the Advisory Board, see the web pages of the Ministry of Transport and Communications Finland, at http://www.mintc.fi/ (at the English language portal see Communications – Information Security and Privacy Protection – National Information Security Strategy) [16.3.2006].

made important by the notion, made especially in constitutional law in Finland, that the procedural rules to protect the essential values of private individuals (the constitutional rights), like the demand for democratic decision-making, speaks about the need of governmental regulation and especially the use of parliamentary laws in the protection of constitutional rights such as information security[1137].

Regulatory literature also makes a serious general suggestion that the substantive value concerns should be determined by a particular mode of decision-making; i.e., participation and deliberation[1138]. This implies the political element; the decisions concerning values ought to be made collectively. For those educated in law this seems to be a familiar solution. Traditionally, and especially when looking law from the inside as legal scholars do, this kind of democratic decision-making is present in the legislative drafting[1139]. More generally, this is what

---

[1137] The basis for this approach is section 80 subsection 1 of the Constitution of Finland (731/1999) concerning delegation of legislative powers according to which "…the principles governing the rights and obligations of private individuals and other matters that under this Constitution are of legislative nature shall be governed by Acts", i.e., laws issued by the parliament (Parliamentary Acts). I do not intend to imply that this provision or the whole Constitution should be interpreted in a manner that necessitates the use of Parliamentary Acts in the regulation of information security. Instead, I only point out that the legislative culture in Finland tends to overemphasize the need to regulate issues at the level of Acts of Parliament. This critique is raised also by the OECD in its review of regulatory reform in Finland from 2003, *Government Capacity to Assure High Quality Regulation in Finland*, p. 5-6 and 29-30.

[1138] See, e.g., Black, Proceduralizing Regulation: Part I, p. 598-599. However, I am not going deeper into the form the participation should take. For further discussion, see the effort of Black in Proceduralizing Regulation: Part I and Part II. Note that this type of proceduralisation is not a distinctive type of moral theory of regulation because it presupposes a prior and independent theory of the moral criteria applicable to regulatory choices (Adler, Beyond Efficiency and Procedure, p. 268-269). True moral theories of regulation give intrinsic, not just instrumental, significance to the fact that the regulator followed or failed to follow some specified type of procedure (a decision-making process). These are what Adler in Beyond Efficiency and Procedure, p. 267, refers to as the proceduralist theories of regulation.

[1139] There is a research and practice area in its own right called legislative drafting. From the Finnish point of view, see Niemivuo, *Kansallinen*

parliaments and governments are for. So, in the case of the law, as an instrument of regulation, and other direct state-centric regulatory instruments, this democratisation is present; laws are formulated in democratic processes where participation and deliberation are inside the process.

There are also other constitutional mechanisms that can be used to restrict the excessive instrumentalism of the 'centred' regulators (i.e., parliament, government, ministries, regulatory agencies etc.) leading to the disconnection from the value base. A central mechanism is the necessary self-restraint of the law put into practice mainly through basic (constitutional) rights control[1140]. In the background of the constitutions of democratic states there is an assumption of efficiency of the state machinery to which the constitution sets limits. Central limitations to the power of this machinery is set by the basic rights provisions; basic rights and liberties, stated not just in national constitutions but also in international conventions on human right, are prerequisites for democracy that have to be protected also from the abuses of power by the state[1141].

From the internal point of view (legal scholarship), the legislative process (providing legally based regulation) is a combination of

---

*lainvalmistelu*.

[1140] This is an argument made by Tuori in *Kriittinen oikeuspositivismi*, p. 234 and repeated by him in Tuomarivaltio, p. 916. Another control mechanism is the separation of powers between legislature, executive and judiciary.

[1141] This does not mean that basic rights and democratic decision-making are in conflict; instead, they are complementary, as pointed out in the Finnish constitutional discussion by Tuori in *Oikeus, valta ja demokratia*, p. 269-281 and by Viljanen in *Perusoikeuksien rajoitusedellytykset*, p. 357. A purely formal conception of democracy will not be sufficient in any event, as a society where the decision-making procedures are democratic in the formal sense, but where the basic rights of individuals are not respected, cannot be considered to be a democracy in any meaningful sense. And at the same time, a functioning democratic decision-making procedure is usually a prerequisite for the true realisation of basic rights. However, basic rights as written in constitutions are not absolute in the sense that they could not be restricted in any way, manner or form, at least in general terms.

political and legal practices[1142]. The political argumentation follows a goal oriented model where legislation is a means to achieve societal goals. Even though this is an instrumental view of legislation (as a tool) it still manages, with the help of constitutional practices, to add the checks of consistence and coherence of the legal order into the legislative process. For example, the content of legislation cannot conflict with the constitution (especially not with the procedures followed in legislating, but also not with the basic rights and liberties) or with the international conventions. The use of legal regulation may be limited also in hierarchical level; at the presence of basic rights and liberties, regulation, when essentially concerning the legal status of a person (her rights and obligations), has to be at the level of parliamentary law[1143].

This is build into the constitutional systems and theories of, at least, western constitutional democracies – the rule of law states. This is the way law operates for those who look it from the inside; it is not just a way of regulating the behaviour of others, it is also a constraint on the behaviour of the regulator. In western constitutional democracies the control of the realisation of basic rights is the main tool used in this.

---

[1142] Tuori, *Kriittinen oikeuspositivismi*, p. 150-153. See also Tala, *Lakien vaikutukset*, p. 2.

[1143] Naturally countries differ in this. Finland is similar, for example, to Germany and United Kingdom where the proportion of Acts of Parliament is high as pointed out by Harmathy in The Influence of Legal System on Models of Implementation of Economic Policy, p. 256-7. However, the reliance in primary legislation that is very specific and detailed command and control type, is extremely high in Finland due to the strong legalistic tradition. And the choice of instruments is even further limited by the new constitution (731/1999) which requires many matters to be regulated by parliamentary acts rather than secondary legislation or other instruments, for example, issues concerning basic rights and liberties, as pointed out also by the OECD in its review of regulatory reform in Finland, *Government Capacity to Assure High Quality Regulation in Finland*, p. 30. In Finland, the traditional view is that basic rights function as a check on the competence of the legislature. But this is no longer realised only in procedural tests of forms of regulation, but also material guarantees of basic rights are considered significant. About the Finnish discussion see, e.g., Viljanen, *Perusoikeuksien rajoitusedellytykset*.

I am not going to argue contrary. Instead, the effort is to widen the ways to protect, and the regulators who can protect fundamental rights. In the light of the decentring and pluralism theses adopted in this study, a requirement of the use of parliamentary laws in the protection of constitutional rights need a re-evaluation.

At first hand, the above considerations seem to speak in favour of state-centred actors, legal instruments and constitutional protection when deciding on the use of regulatory tools for the enhancement or protection of certain values in a situation requiring regulation. It is easy to make the excessively restrictive notice that the participatory and deliberative processes needed in deciding about the substantive values protected in certain regulation are present only within the legislative drafting. In other decision-making processes the participatory and deliberatory functions easily just does not seem to be present in a sufficient degree. And that the decision-making about values should be collective, which is the role of governments.

It is equally easy to limit the sufficient tools for the protection of basic rights to those of parliamentary laws: for many (of those educated in law) the requirement of the constitutional restriction of regulation - constitutional protection from the misuse of power by the regulator – calls out the use of governmental regulation and especially parliamentary laws (or regulatory powers deriving from parliamentary laws). The correct desired values held by the population in question just seems not to be protectable otherwise.

Despite the correctness of the arguments that seem to favour governmental regulation, they are not adequate. They lack in scope. I see at least five reasons for this.

*Firstly*. Deliberation, as a means for deciding about the substantive values enhanced by regulation, need not be restricted to legislature and the courts, not even if introduced into the administration as Habermas seems to desire[1144]. There is no need to undertake the inadequate conception of deliberative democracy existing only within a constitutional state (state following the rule of law). Deliberation

---

[1144] Habermas, *Between Facts and Norm*, p. 426, 440-442.

and participation as the basis for norm formation can occur also outside the state apparatus.

This is actually what Julia Black argues by developing the Habermasian discourse principle into the basis of norm formation outside the legislature, courts and administration; deliberation in accordance with the discourse principle but involving a fragmented state[1145]. She deviates from the Habermasian implication that deliberation in accordance with the discourse principle occurs only in the legislature and the courts (and ought to be introduced into the administration). For Black, this is too great a restriction on the potential for other loci of deliberation. This is also what Michael Froomkin implies when he suggest that complex non-governmental international rulemaking discourse conducted by the Internet Engineering Task Force (IETF) is a concrete example of a rulemaking process that meets Habermas' notoriously demanding procedural conditions for a discourse capable of legitimating its outcomes[1146].

*Secondly*. Choices made about value-laden regulatory tools need not always be political decisions that should necessarily be subject to collective decision-making inside the state. In relation to the indirect instruments, especially social norms or technology harnessing and market-based instruments, the same choices between different regulatory tools and the embedded values can also be an aggregate outcome of uncoerced individual decisions.

This may even be a better way. Without a central plan, i.e., a collectively decided public policy of the tools, there is a more diverse set of offerings before the members of the public in response to their diverse needs and preferences. The values can be protected by allowing the widest possible scope for uncoordinated and uncoerced individual choice among different values and among different embodiments of those values[1147].

---

[1145] Black, Proceduralizing Regulation: Part II, p. 35-36.

[1146] Froomkin, Habermas@Discourse.net, p. 752 and 754-755.

[1147] This is the argument of David G. Post in his critique of Lessig's *Code and Other Law of Cyberspace*, What Larry Doesn't Get, p. 1450-1459, in relation to the values in the architecture in the Lessigian sense. Post uses the example of

'Can' is enough; one does not have to subscribe entirely to this argument. The possibility of other ways to make the decisions concerning values embedded in regulatory tools is sufficient to straighten the monopolistic conception that *collective* decision-making, together with participation and deliberation, is needed to make decisions about values and regulatory tools that embed those values.

*Thirdly*. Neither are the necessary self-restraints for regulation present in constitutional systems applicable only to parliamentary laws. The constitutional constraints on regulation apply also to the indirect forms. When a direct constraint on behaviour (e.g., through law) conflicts with basic rights or with other self-restraints of basic rights control, it raises concerns whether that same constraint should be allowed to be effected indirectly through the market, social norms, or technology[1148]. Even though the question of constitutional self-restraints applying to indirect forms of regulation is somewhat uncertain and has not been widely discussed, there is a logical appeal in the argument. Why should regulators be able to circumvent the constitutional protections for basic rights and liberties merely by using indirect regulatory means?

*Fourthly*. Neither do the basic right provisions only protect individuals and groups from government interference. The constitutional constraints on regulation work also towards other regulators than just the legislature. This can be derived from the role of basic rights and liberties as subjective rights that express fundamental value decisions in society and as legal principles that require optimization. Basic rights no longer express the fundamental

---

the development of language and the role of collective decision-making in that process. This is actually also the message of welfare economics studies in regulation. Actors in a 'market' should be left free to make decisions, unless the decision-making process is disturbed in a way that prevents the socially best outcomes. Naturally, under the influence of network effects both the provision of cultural values and the different market for technologies can lead to winner-take-it-all situations and customer lock-in (thus leading to near monopolies), but it necessarily does not have to do so. The relevance of collective protection of values becomes relevant only when there is a real possibility of this.

[1148] Lessig, The New Chicago School, p. 688.

order of state or the public authorities, but the basic order of the whole society. The argumentation involves two trends in current basic rights theory: horizontal effects and obligation to secure realisation.

First of all, as expressions of societal value decisions and as legal principles basic rights no longer have an effect only on the vertical relationship between the individual and the state, but also on the horizontal axis between individuals and the whole society. Current constitutional doctrine acknowledges the horizontal effect of basic rights and liberties; they protect individuals and groups not only from governmental interference but also from the interference of other individuals or groups[1149]. In addition to having force in the interpretation and application of parliamentary laws, the basic rights provisions also are directly applicable norms.

Secondly, as expressions of societal value decisions and as legal principles basic rights require realisation: the public authorities[1150] have an obligation to secure the realisation of basic rights and liberties and human rights as acknowledged in current constitutional doctrine and explicitly laid down in the Section 22 of the renewed Constitution of Finland (731/1999)[1151]. The legislature is thus constitutionally obliged to protect one group of private actors from other private actors' intrusions on their freedom.

NSAs as regulators do more than a merely exercise individual freedom; they also do more than simply exercise economic or social power. A private governing body does not act purely in its own interests; it has an impact on the freedoms of others in the interest of third party (e.g., in the case of lawyers association this third party

---

[1149] Of special problems in the legal systems in USA see, Berman, Cyberspace and the State-Action Debate.

[1150] Mainly the legislature due to the principles of democracy and distribution of powers. Tuori, Tuomarivaltio, p. 929 and 939.

[1151] Section 22, titled Protection of basic rights and liberties, states that the public authorities shall guarantee the observance of basic rights and liberties and human rights. The Constitution of Finland is available in English from the web pages of the Ministry of Justice at http://www.om.fi/uploads/54begu60narbnv_1.pdf [22.2.2006]. The distinction between basic rights and their guarantees is emphasized by Luigi Ferrejoli in Fundamental Rights, p. 23.

would be the customers or clients of the lawyers that are members of the association). When NSAs regulate behaviour, they do not only enjoy the protections of the constitution, but they also are subject to the constraints in basic rights systems towards their regulatees.

*Fifthly*. Even if bringing the regulation by NSAs, networks or hybrids under the self-restraints in basic rights provisions makes sense, the use of basic rights control is not the only possible way to restrain the use of power. It is not necessary to rely solely on doctrine of basic rights as expressions of societal values or as legal principles. Cultural values in relation to social norms operate in a similar way; they delineate the constraints for and give meaning to individual behaviour. The use of social norms in regulation is constrained by the general social environment that gives meaning to actions, defines what is socially acceptable, and exercises social control through sanctioning. The social environment is shaped by the values individuals have adopted because, to the extent individuals have acquired culture, they can predict how their actions would likely to be perceived by others in their environment, such as professional peers, family, the media, and so forth.[1152]

In sum, neither the decision-making in relation to substantive values of regulation, nor the protection of constitutional values require the use of parliamentary laws or speak only in favour of state-centric regulation. Questions concerning values can also be determined in other places than just inside the democratic procedures of constitutional states. The protection of those values neither requires that only parliamentary laws are used nor sets a prerequisite for regulation by the state centred actors. The regulation of information security does not have to be based, at least not solely, on parliamentary acts. No matter how we systematise information security among constitutional rights, which is largely a matter of controversy, there is no need to rely purely on parliamentary acts in the regulation of information security issues.

---

[1152] This argument has been developed, e.g., by Amir N. Licht (2005) in Social Norms and the Law: A Social Institutional Approach, p. 36, Interdisciplinary Center Herzliya, Radzyner School of Law, Working Paper , available at http://www.faculty.idc.ac.il/licht/PSN7.4-SSRN.pdf [22.2.2006]

## Name Index