Aspects of internet security - identity management and online child protection

Aspects of Internet Security – Identity Management and Online Child Protection

By
Chris Durbin

A doctoral thesis submitted in partial fulfilment of the requirements for the award of Engineering Doctorate at Loughborough University

December 2010

Abstract

This thesis examines four main subjects; consumer federated Internet Identity Management (IdM), text analysis to detect grooming in Internet chat, a system for using steganographed emoticons as ‘digital fingerprints’ in instant messaging and a systems analysis of online child protection.

The Internet was never designed to support an identity framework. The current username / password model does not scale well and with an ever increasing number of sites and services users are suffering from password fatigue and using insecure practises such as using the same password across websites. In addition users are supplying personal information to vast number of sites and services with little, if any control over how that information is used.

A new identity metasystem promises to bring federated identity, which has found success in the enterprise to the consumer, placing the user in control and limiting the disclosure of personal information. This thesis argues though technical feasible no business model exists to support consumer IdM and without a major change in Internet culture such as a breakdown in trust and security a new identity metasystem will not be realised.

Is it possible to detect grooming or potential grooming from a statistical examination of Internet chat messages? Using techniques from speaker verification can grooming relationships be detected? Can this approach improve on the leading text analysis technique – Bayesian trigram analysis? Using a novel feature extraction technique and Gaussian Mixture Models (GMM) to detect potential grooming proved to be unreliable. Even with the benefit of extensive tuning the author doubts the technique would match or improve upon Bayesian analysis. Around 80% of child grooming is blatant with the groomer disguising neither their age nor sexual intent. Experiments conducted with Bayesian trigram analysis suggest this could be reliably detected, detecting the subtle, devious remaining 20% is considerably harder and reliable detection is questionable especially in systems using teenagers (the most at risk group).

Observations of the MSN Messenger service and protocol lead the author to discover a method by which to leave digitally verifiable files on the computer of anyone who chats with a child by exploiting the custom emoticon feature. By employing techniques from steganography these custom emoticons can be made to appear innocuous. Finding and removing custom emoticons is a non-trivial matter and they cannot be easily spoofed. Identification is performed by examining the emoticon (file) hashes. If an emoticon is recovered e.g. in the course of an investigation it can be hashed and the hashed compared against a database of registered users and used to support non-repudiation and confirm if an individual has indeed been chatting with a child.

Online child protection has been described as a classic systems problem[1]. It covers a broad range of complex, and sometimes difficult to research issues including technology, sociology, psychology and law, and affects directly or indirectly the majority of the UK population. Yet despite this the problem and the challenges are poorly understood, thanks in no small part to mawkish attitudes and alarmist media coverage. Here the problem is examined holistically; how children use technology, what the risks are, and how they can best be protected – based not on idealism, but on the known behaviours of children. The overall protection message is often confused and unrealistic, leaving parents and children ill prepared to protect themselves. Technology does have a place in protecting children, but this is secondary to a strong and understanding parent/child relationship and education, both of the child and parent.

[1] S. Brown (INCOSE President Elect), Personal communication, 21st April, 2009.

Acknowledgments

First and foremost I must mention *Mark Pawlewski* my industrial supervisor without whose tireless patience, support, and encouragement this would never have been completed.

I also wish to thank my academic supervisor *Prof. David Parish* for his encouragement and guidance.

I should also mention *Roger Payne* who served frequently as a sounding board.

Finally, I wish to thank my family for their belief and support.

Table of Contents

1. Introduction 11
1.1 Identity Management 12
1.2 Online Child Protection 13
1.3 Thesis Structure 15
2. Identity Management 17
2.1 Circle of Trust and nomenclature 18
2.2 Features and Goals for a Successful Identity Metasystem 19
2.3 Internet Revolutions 20
2.4 No Business Model 21
2.5 The Identity Management J-Curve 22
2.6 Moving Towards a New IdM System 24
2.7 Technologies of Interest 25
2.8 Conclusion 26
2.8.1 Facebook as an IdP 26
2.8.2 Government Intervention 28
2.9 Business Impact 28
2.10 Identity Paper 30
3. Text Analysis 41
3.1 Text based Internet conversations 42
3.2 Automatic Grooming Detection 42
3.3 Classification Tests 43
3.3.1 Simple trigram 44
3.3.2 Bayesian 44
3.3.3 Gaussian Mixture Model (GMM) 46
3.4 Results 48
3.5 Discussion 48
3.6 Conclusion 49
3.7 Business Impact 50
3.8 Text Analysis and Message Review Paper 51
4. Steganographed Custom Emoticons 60
4.1 Background 61

4.2 Catching Internet Criminals 61
4.3 MSN 62
4.4 ‘Steganographing’ and Detecting Custom Emoticons 64
4.5 Method, Experimentation and Conditions 65
4.6 Justification and Discussion 68
4.7 Conclusion 71
4.8 Business Impact 71
4.9 Steganographed Emoticon Report 73
5. The Online Child Protection Domain 85
5.1 Systems Engineering thinking applied to the problem 86
5.2 What is paedophilia? 87
5.3 Risks to children online 90
5.3.1 Bullying 91
5.3.2 Unwanted Exposure 92
5.3.3 Solicitation 94
5.4 Threat vectors 95
5.5 Social Networking 97
5.6 Child Pornography 102
5.7 Current safety / prevention message 106
5.8 Examination of current protection software 107
5.8.1 Content Filtering 107
5.8.2 Monitoring 108
5.8.3 Moderation and Pre-moderation and Wall Gardens 109
5.8.4 Network Level Censorship - IWF and CleanFeed 110
5.8.5 Community Reporting and ‘Safety Mode’ – Learning from YouTube 111
5.9 Security Theatre 112
5.10 Realigning risk 114
5.11 Trade offs and Best use of Resources 116
5.12 Difficulties Posed to Researchers 118
5.13 Discussion 120
5.14 Conclusions and recommendation 120
5.15 Business Impact 122
5.16 Reflective Commentary 123
5.17 Systems Analysis Report 125

6. Conclusions and Further Work 141
6.1 Identity Management 141
6.2 Text Analysis 143
6.3 Steganographed Custom Emoticons 144
6.4 Online Child Protection 147
7. references 151
7.1 Identity Management 151
7.2 Text Analysis 154
7.3 Steganographed Custom Emoticons 155
7.4 Child Protection 157
8. Appendix A – Terminology 171
9. Appendix B – Risks and safety on the internet 173
10. Appendix c – Flowcharts 174
10.1 Text Analysis 174
10.2 Steganographed Custom Emotions 177
11. Appendix d - Custom Emoticon Change – Experiment Log 179

Table of Figures

Figure 1 - Thesis Structure 12
Figure 2 - Circle of Trust 18
Figure 3 - The IdM J-Curve 23
Figure 4 - Identity Theft Complaints (Source FTC [5]) 24
Figure 5 - Trigrams 44
Figure 6 - Feature Vectors (Underscores represent spaces) 47
Figure 7 - Vectors representing frequency of occurrence 47
Figure 8 - Default vs custom emoticon transfer 63
Figure 9 - Source and steganographed emoticons and hashes 65
Figure 10 - Emoticon manufacture and issue 66
Figure 11 - Emoticon recovery process 67
Figure 12 - Sex offender motivation continuum (Lanning) 89

Table of Tables

Table 1 - Classification results matrix 48
Table 2 - Minors solicited or harassed online in the past 12 months, (Source [41]). .. 97
Table 3 - COPINE Scale (CS) and SAP = Sentencing Advisory Panel (SAP) 103

1. INTRODUCTION

This thesis represents the results of work performed during a four year placement at BT Research Labs. The work started with a broad investigation into Identity Management (IdM) resulting in a published discussion paper [9] examining the obstacles to Identity Management becoming established in the Retail Sector, where the end users are general members of the public. The main finding of this investigation was that that the general public are unlikely to adopt IDM technologies in the foreseeable future unless there is immediate tangible benefit and user friendly implementation. After this investigation of the IDM space, the work moved toward looking at the general problem of online child protection. BTs initial requirement was to look at technologies that purport to protect children from on line predators who may attempt to befriend them in chat rooms or via MSN, with illegal underlying intentions; otherwise known as Child Grooming. The initial work looked at technologies for detecting the difference between text chat from a child and text chat from an adult who is masquerading as a child. There are several issues around testing such technologies in that appropriate test data is not easily available. Nevertheless, despite the problem with data availability, in order to get a sense of the possible accuracy of such systems, an analysis of text from different authors was performed. The aim here was to establish if an appropriately trained classifier could distinguish between the texts of various authors. The degree of difficulty for this task was, reasonably, assumed to be the same as that of identifying written text of adults pretending to be children, compared to that of genuine children. Although initial results looked promising, the research established that it is indeed very difficult to spot subtle differences between author styles and from this we can infer that the same is true for separating chid text chat from pseudo child text chat. This research then led on to two main strands of work, the first looked at novel methods for protecting children on line, culminating in the development of a technique for legally placing a digital ‘fingerprint’ on the machine of a potential groomer, thus proving that there has been interaction between their machine and the machine of a child running the new protection software. The second strand of work concentrated on looking at the larger picture of online child protection and basically concluded that technology only plays a very small role in child protection. It is much more important to understand the circumstances under which children come under threat and the best ways to mitigate

these threats. This last area of work is arguably the most useful to BT in that it provides a deeper understanding of the problem which is not immediately evident to the uninformed.

The success of this work has led the author to become a consultant for BT advising on Child Protection issues.

Figure 1 illustrates the structure of this thesis and how it relates to the larger field of Internet security.

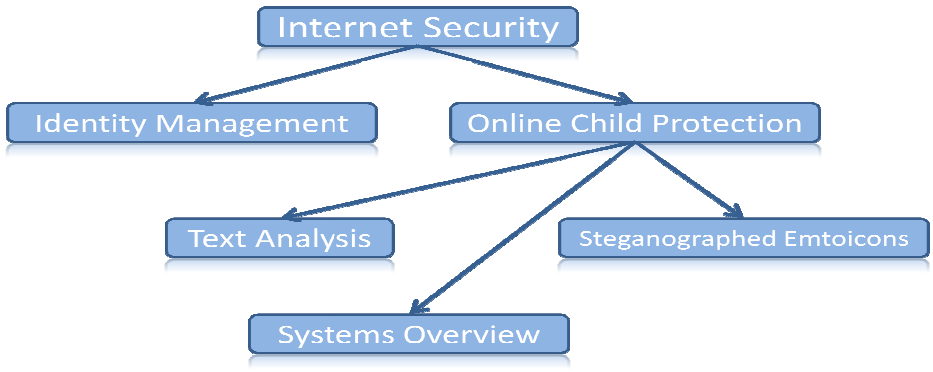


Figure 1 - Thesis Structure

1.1 Identity Management

Users are familiar with sites that require them to sign up to websites in order to gain access to services and buy products. The vast majority of these sites record the users' personal details and issue them with a username and password for use with that site only, or perhaps other sites operated by the same organisation. Users are becoming overwhelmed with a plethora of online identities which has lead to password fatigue - using the same username and password across multiple identities and recording passwords in plain text.

In addition to the problem of password fatigue users are revealing a great deal of personal information when registering for services and products. This information is

often unnecessary for the provision of services and products and increases the likelihood of such information falling into criminal hands and the user becoming a victim of identity theft.

Identity management is common in the enterprise where central policy can dictate how and where identity is used, as well as bearing the associated costs. In the unmanaged consumer domain the future for identity management is much less clear.

A feasibility study of consumer Internet identity management and federation revealed that while no insurmountable technical challenges exist, user interest, user experience and business model are a significant barrier to a new identity metasystem. In the short to medium term there will be no significant changes to the consumer Internet identity management space. Users will continue to reveal personal details when signing up to a new product or service. Products and services will continue to identify users with usernames and passwords. More security conscience platforms, typically those providing financial services will increasingly use one time passwords and other hardware tokens. Without a severe breakdown in Internet trust and security, this status quo is set to remain. The author has presented a conference and journal paper [9] detailing this study.

1.2 Online Child Protection

The major theme of the author's research was online child protection - this began as a project to see if child groomers can be identified from the contents of online messages. During this work the author reviewed the literature in order to gain a broader understanding of the problem, this lead to an expansion of the work into a systems driven meta-analysis of the problem space. During the ongoing meta-analysis a side project was completed that allowed 'digital fingerprints' to be left on the computer of individuals who chats with a child on the *MSN .NET Messenger Service*.

Is it possible to detect when a child may be chatting to a groomer based on the content of the conversation? No longer can parents see who their children are chatting with, this makes keeping them safe increasingly onerous. Children often conduct online friendships in chat rooms, by instant message, email and social networks. These conversations can be in real time or time lapsed. Research has evaluated the feasibility

of detecting potential grooming by conducting statistical analysis on the features of text based conversations. A novel technique of feature extraction and classification - building 'sliding windows' of character strings and Gaussian Mixture Model (GMM) is described and tested against the leading text classifier – Bayesian trigram analysis.

Experiments revealed that obviously differing text, for example sexual explicit conversation as might be found in grooming could be reliably detected. More subtle, long term grooming is much harder to detect and the reliability of classifiers drops accordingly. Ultimately the novel feature extraction and GMM classifier proved unsuccessful and was inferior to the current leading technique.

With ever more children enjoying online friendships the possibility of being groomed also increases. A method has been devised to create a 'digital fingerprint' by manufacturing a unique emoticon using steganographic principles. This digital fingerprint can be left on the computer of anyone who chats with a child by exploiting properties of the *MSN .NET Messenger Service.* This works as a transparent background function and involves no hacking. At some later time the emoticon/fingerprint can be recovered, for example during a law enforcement investigation and used to prove the recipient engaged the child in conversation. A function of the chat clients changes the file contents if the emoticon is reused by the recipient eliminating false positives.

The meta-analysis examined the online child protection domain from a systems perspective. This indicated that perception of the issues and parental fear is largely disproportionate to reality, driven to a significant degree by media and political interest. The analysis revealed that the dangers facing children are essentially the same they have always faced; only the delivery agent, the Internet, has changed.

Solicitation, bullying and exposure to unwanted material are the main risks facing children. Solicitation is rare and children handle it well, similarly the majority of children were unfazed when exposed to unwanted material. Bullying is not uncommon and may be a source of long term emotional distress.

Technology does have a part to play in protecting children, especially younger naïve children; it is much less important where older, head strong, technologically competent teenagers are concerned. Social measures including education and family relationships are more powerful tools for protecting older children.

A significant factor retarding progress in the field is a lack of agreed definitions, standards and metrics which prevents any side-by-side comparison of data and means scientifically rigorous work on a subject can produce widely differing results.

1.3 Thesis Structure

Chapter 2 examines the existing situation, the requirements of a new system and high level perspective of how such a system would work. It discusses two possible futures for consumer identity management and the conditions each would require and the barriers to a new system.

Chapter 3 describes the text analysis work; it investigates the various approaches to feature extraction, entropy handling and classifiers performance and suitability to task. It also describes the novel feature extraction and GMM classifier and the experiments against the current techniques.

Chapter 4 details the steganographed emoticon / 'digital fingerprint' project; it discusses the birth of the emoticon, how the emoticon is manufactured, the operation of *.NET Messenger Service*, the use of hashes to identify the emoticon and the experimental work.

Chapter 5 describes the Internet child safety meta-analysis; it uses the systems approach to examine the problem space from a holistic perspective. It examines the complexity inherent in a social problem of this size, it looks at stakeholders, technologies, human factors and behaviours, media and political involvement, the risks facing children and how these can best be addressed.

Chapter 6 recaps the key points from each chapter and discusses the conclusions drawn from the findings.

2. IDENTITY MANAGEMENT

The explosion of online applications, services and websites has lead to an equally large plethora of credentials – usernames, passwords and one time password (OTP) key fobs and card readers. With few exceptions each website, service or application requires separate registration, username and password. This multiplicity of accounts has led to 'password fatigue' - insecure practices such as writing down passwords in clear text and using the same username and password for multiple accounts. In addition users are supplying personal information to numerous organisations (the website and service operators), often with little knowledge of how secure, or trustworthy these organisations are.

In the real world identity documents often assert specific claims about the user, for example being over 18, or possessing a driving licence. The traditional username / password approach of Internet security lacks the functionality necessary for users to prove such assertions.

There has been much hype and evangelising in the blogging community [1] [2] about the coming identity revolution. Often this refers to a federated identity metasystem[2], where users possess a portable (use anywhere) identity, are able to assert specific claims and control the disclosure of personal information. At present there has been only limited realisation of this.

At the root of the Identity Management[3] (IdM) challenge lays the design and structure of the Internet. It was not designed to support digital identities; the current approach of username / password has evolved organically over time to meet the needs of a single provider (website or service) but is inadequate to meet the needs of the modern Internet. Attempts have been made to address these problems – most notably *Microsoft Passport*[4] - but, for a variety of reasons, they have met with failure – the lack of provider and consumer buy-in, a closed proprietary approach, inadequate user

[2] An interoperable architecture based upon multiple underlying technologies and providers.
[3] Defined as an individual's ability to manage their identity as they join, leave and interact with an organisation.
[4] https://accountservices.passport.net/ppnetworkhome.srf?vv=800&mkt=EN-GB&lc=2057 (accessed on 23 June 2010).

controls [3], lack of privacy [17], data loss (expose) and a lack of security of the data silo [18] and the lack of a decentralised architecture [19].

A new breed of IdM systems promises to meet the needs of the modern Internet by providing the existing functionality of logging-in but also being able to verify individual claims about the individual (age for example)and be portable across domains (different websites and services)and place the user in control of disclosure.

2.1 Circle of Trust and nomenclature

Although the language, implementation and design goals of the various systems differ they all employ the notion of a Circle of Trust (see figure 2). In identity parlance the Identity Provider (IdP) is the repository for user information and personal details. A Reliant Party (RP) is any platform, service, website or application that a user would currently be required to login to. Rather than each RP being an isolated silo of user information, the identification process (of consumer to RP) is mediated by an IdP – the IdP testifying to the user's identity and providing any credentials the RP requires.

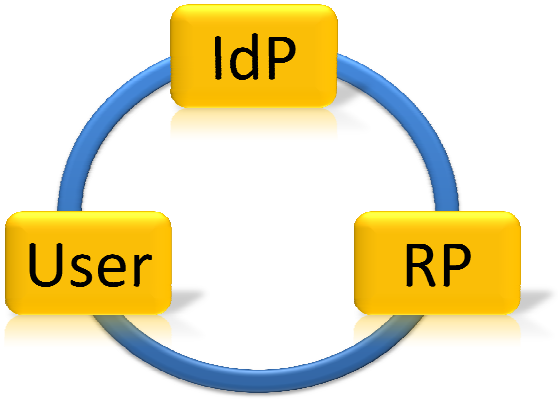


Figure 2 - Circle of Trust

The need for a new identity system is not in doubt, but users are not ready to embrace it [4]. At present only a handful of IdPs exist on the Internet and they provide very limited functionality. There are no signs now, or in the short term future, that an IdP will emerge to offer non-trivial functionality, such as online retail or online banking to consumers.

This chapter of the thesis will examine what is necessary for a workable identity metasystem to emerge, what the common goals are and why this will not happen without a culture change among users in the Internet community. IdM has made successful inroads in the enterprise with products from the likes of *Microsoft*[5] and *Oracle*[6] offering identity management within and between organisations. The enterprise controls and administers the identity of its employees (within the scope of their employment); it also controls the RPs, the various systems employees need to access. It does not need user buy-in or support from RPs. Unlike the democratic nature of the Internet, the enterprise is able to mandate IdM with carte blanche, this accounts for the success of enterprise IdM. In this chapter the focus will be on consumer IdM, that is, the management of individuals' identity as they use the Internet.

2.2 Features and Goals for a Successful Identity Metasystem

To avoid the failings of *Passport* a successful identity metasystem will have to follow common goals and features.

- Federation – Users need one identity (or a limited number) and use this to provide their identity anywhere with Single Sign On (SSO) capability, much as an individual can use a credit card anywhere. There will likely need to be an existing relationship between the IdPs and RPs.

- Login Security – Passwords are often a weak link; weak passwords are at risk of cracking, complex passwords are often written down by users. Multi-factor authentication, in the form of hardware based or mobile phone OTP are suitable for this purpose.

- Information Security – Identity theft has been on the rise in recent years [5], for an IdP to gain traction users must be confident their IdP can protect their personal information.

[5] http://download.microsoft.com/download/3/a/f/3af89d13-4ef4-42bb-aaa3-95e06721b062/ADFS.doc (accessed 09 June 2010).
[6] http://www.oracle.com/technology/products/id_mgmt/index.html (access 09 June 2010).

- User Centricity – Consensus suggests that users want to be in control of how and when their identity is used [6] and should be able to withdraw from a transaction if an RP requests excessive information. Users must be able to control how information is released in order to maintain privacy, for example when using adult services, online gaming or simply separating their work and private lives [7] [8].

- Standards – To ensure interoperability between IdPs and RPs a common standard or suite of standards is necessary, without which the system is unlikely to be adopted by users.

2.3 Internet Revolutions

Martin, Durbin, Pawlewski and Parish [9] argue that IdM will not see a sudden surge in user uptake typically seen in 'Internet revolutions' such as email, social networks and VoIP. Each of these 'Internet revolutions' is very different to what came before, but has a *real world* counterpart (mail, offline social networks and telephones) but the Internet introduced a key selling point[7] – added functionality, speed, convenience, accessibility or price. The technology itself isn't necessarily revolutionary, the emphasis is on user experience and how it changed the nature of user interaction.

Real world revolutions are driven by grass roots individuals, rapidly gain support and usurp the previous administration. Similarly 'Internet revolutions' quickly gain in popularity among users and replace (or compliment) what came before.

Increased or improved security adds no 'useable' feature that users can enjoy – it does not effect the user experience, it hardens (the security) that already exists. Individuals are frequently concerned with security only once it has failed, this can be observed in the mindset 'it does the job', indeed users may not be aware of, or understand the issues and risks surrounding security [15] [16]. This has inhibited a grass roots interest among consumers which seems unlikely to change without a cultural shift in attitudes.

[7] 'Selling' in terms of selling an idea or concept, rather than a transaction for goods or services.

2.4 No Business Model

The long term success for any organisation is dependent on its business model, essentially how it generates funding. Numerous business models exist including selling products or services, advertising, external revenue e.g. charitable donations or subsidy. Unsustainable business models were responsible for the failure of many organisations in the Internet 'dot-com bubble' [102].

There are no IdPs that support financial transactions and none operate with financial institutions (banks, credit card companies). At present IdPs only support simple, non-financial applications such as blogging. The general public is resistant to paying for services on the Internet (particularly those they currently receive for free) and there is no incentive for RPs (retailers, service providers, banks, etc) to fund IdPs.

Customer records and transaction histories are a rich source of marketing information providing retailers and service providers opportunities for upselling and cross-selling. Some organisations even sell customer records and mailing lists to marketing organisations. Surrendering this information to IdPs, reducing to a minimum amount the customer data they held, would curtail a potentially lucrative revenue stream for RPs.

When a fraudulent transaction takes place the credit card company or bank, in the absence of negligence (on the part of the consumer), absorbs the loss as an operating cost and refunds the consumer. If a financial loss was incurred as a result of failings at an IdP, the IdP would likely be liable. If a fraudster managed to comprise an individual's account they could conceivably have access to all the user's financial details (current account, savings, mortgage, shares, and credit cards). If a fraudster successful executes a series of attacks or worse, a class attack[8] - the liability exposure could run into millions of pounds. In 2007 retailing umbrella group *TJX* (owners of , among others *TJ Maxx*) agreed to pay banks $40.9M after up to 100 million credit numbers were sniffed (intercepted) from insecure networks [10].

[8] An attack against a class of products or services, e.g. a vulnerability in Windows XP.

With neither customers nor RPs willing to providing a revenue stream and the potential of enormous liability exposure, the business model appears unsustainable.

2.5 The Identity Management J-Curve

Given the status quo, the appearance of a non-trivial identity metasystem seems improbable. Could a catastrophic breakdown in trust and security on the Internet (with associated financial losses) force the issue to the fore and provide the required impetus?

Consumers are resistant to paying for IdPs as a solution already exists and levels of identity theft are low. Millions of people in the UK have seen their personal information exposed, either through identity theft or data mismanagement [11], but the consequences have been few. As a percentage of the total number of records lost, the number of individuals suffering financial loss or credit rating difficulties is low. Without this financial penalty much apathy surrounding IdM will likely remain.

The J-Curve is used in economics, political stability, medicine and elsewhere. Bremmer [12] uses the J-Curve to plot the relationship between political stability and government openness – when moving from a closed, secretive state, to an accountable democracy, national stability suffers. Martin *et al* [9] use the J-Curve (see figure 3) to illustrate how a breakdown in Internet trust and security would lead to an open federated identity ecosystem i.e. things will have to get worse before they get better.

Moving to the left of the J-Curve would improve identity security and stability through the hardening of information silos and processes (proprietary protocols, complex passwords, OTP devices). This would do nothing to address the underlying issues of IdM and security improvements might prove temporary until Internet criminals discovered and exploited new vulnerabilities.

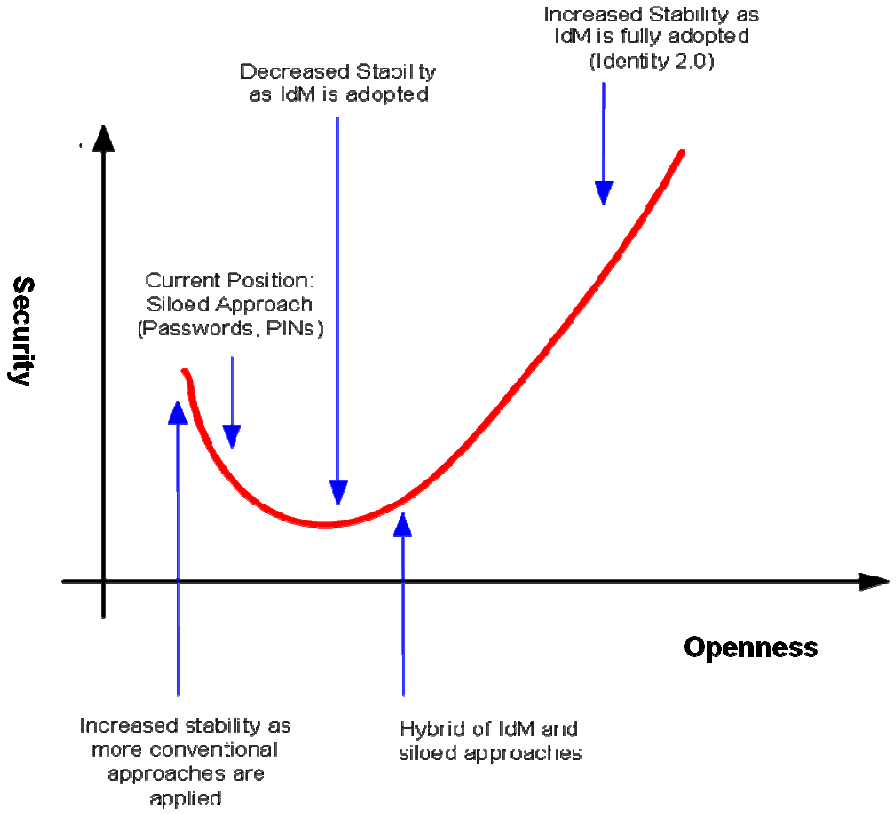


Figure 3 - The IdM J-Curve

A move to the right would see the rise of IdP and RPs operating in a federated infrastructure based on open standards. Initially this would see a reduction in security as the ecosystem suffered 'teething problems' (as a result of improper configurations, user confusion etc). Over time the ecosystem would 'settle down' and security would improve – the more consumers, IdPs and RPs adopting the federated IdM approach, the more security and stability would improve. In addition to improvements in security and stability the federated IdM approach would also address the underlying failings in the current siloed approach.

The current state of the IdM ecosystem is to the left of the trough, for an immediate improvement in security it would be necessary to move to the left. Indeed this is happening in the online banking world – the launch of 'Chip and PIN' in the UK (though not an online process) is an example of a hardening procedure. Some UK banks have issued OTPs to customers, these are providing security improvements but at the expense of convenience and address none of the larger IdM issues. This is not an ideal state of affairs; more desirable would be a move to the right, with open

protocols, federated infrastructure and user centricity featuring minimal disclosure and a consistent user experience.

2.6 Moving Towards a New IdM System

The current state of IdM is not sustainable. Sources estimate that 25% of UK adults have had their identity stolen or know someone who has [13]. These figures are increasing year-on-year – see figure 4. If these levels and the losses associated with them become endemic this could provide the impetus necessary to see a modern IdM metasystem take hold. As more and more individuals continue to incur losses, there maybe a sea change which finds the public willing to pay for IdPs. Comparisons can be drawn with the uptake of anti-virus, anti-malware, anti-spam and other Internet protection software that was driven not by popularity or novel features, but by the necessity of protecting users and computers on the Internet. These Internet protection products are only popular because of the ubiquity of Internet threats, a considerable rise in identity theft and in particular personally incurred financial losses would be necessary to realise this affect to the IdM ecosystem.

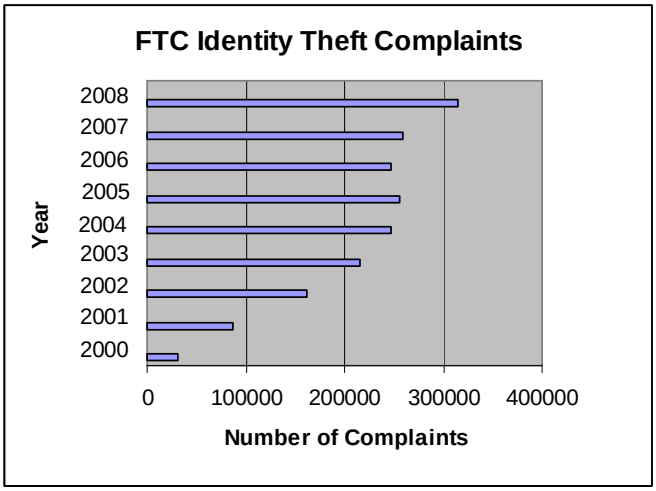


Figure 4 - Identity Theft Complaints (Source FTC [5])

In the absence of consumer driven change in the identity ecosystem, the government could mandate the use of Internet IdPs. This might be feasible within the UK banking sector, but would likely meet limited success in retail and service provision. Any government mandated requirements would only be enforceable in the country of origin and would doubtfully see the universal uptake necessary to reform IdM.

Lack of user education, user understanding (of the larger problem) and user concern are retarding a federated IdM metasystem. Users aren't aware that an alternate approach exists, nor of the advantages it offers. This might provide an opportunity for government intervention, in the form of education and awareness. In the UK the government participates in a joint venture called *Get Safe Online*[9] to provide free advice on Internet safety. If suitable IdM infrastructure did exist, the government could use this platform, or one like it, to assist in user education and promotion.

2.7 Technologies of Interest

Among the new crop of IdM technologies two are particularly worthy of mention - *Microsoft CardSpace*[10] and *OpenID*[11].

CardSpace [14] from *Microsoft* is a cryptographically secure product based on the *InfoCard* standard, developed around the 'Identity Laws[12]' of project champion Kim Cameron. The project is open, extensible and encourages third party contribution. This approach overcomes the failings of *Passport*, but still has the penetration and market share of *Microsoft Windows*. *CardSpace* is a standard feature of the *Vista* and *Windows 7* operating systems and an option for *Windows XP*. Despite this a *CardSpace* infrastructure has yet to emerge.

OpenID is a lightweight browser-driven (requires no additional software) SSO tool, based on standard web protocols. *OpenID* is significant because it *has* been subject to grass roots driven success – it's popular in the blogging community where visitors can comment on blogs without registering, by logging in with their *OpenID*. *OpenID* is a decentralised protocol that requires no existing relationship between IdPs and (*OpenID* enabled) RPs. An inherently insecure protocol *OpenID* is susceptible to 'phishing' attacks and unsuitable for financial transactions. *OpenID* now has support for *CardSpace* hardening it against this type of attack.

[9] http://www.getsafeonline.org/ (accessed on 17 June 2010).
[10] http://www.microsoft.com/windows/products/winfamily/cardspace/default.mspx (accessed on 18 June 2010).
[11] http://openid.net/ (accessed on 18 June 2010).
[12] http://www.identityblog.com/stories/2004/12/09/thelaws.html (accessed on 18 June 2010).

2.8 Conclusion

The need for a modern federated IdM metasystem is clear and though common in the enterprise, it has not found success in the consumer market. There is no business model to support the operation of an IdP, users see it as offering nothing exciting and will not pay and there may be extensive liabilities involved. Businesses will be unwilling to pay and not keen to surrender customer details that offer a wealth of marketing opportunities.

The Identity J-Curve illustrates the current state of consumer Internet IdM, as present society is some way left of the trough and moving further left. This move to the left is being driven by the rise in fraud and identity theft. It has led to a hardening of identity silos, enforcing of complex passwords and proprietary, per application OTP devices. The IdPs that do exist only offer access to low value applications such as blogging.

The challenges are well understood, the technologies and protocols necessary to build a functioning identity metasystem exist today. As long as customers are unwilling to pay, businesses remain resistant to loss of control and no other business model arises the short-to-mid-term outlook for IdM will be a continuation of the status quo. This will be a continuation of more username / passwords, complex passwords, OTPs and hardware devices. Should there be a breakdown in Internet trust and security, this may provide the impetus necessary to seed a modern federated IdM metasystem. If a significant percentage of online transactions are fraudulent and consumers are financially exposed to losses and the inconvenience associated with identity theft, they may be willing to pay for an IdP. This might find support from banks and credit card providers if losses become endemic (indeed, financial institutions would make ideal IdPs). In the event of a serious decline in Internet trust and security a government sponsored push perhaps from the US or EU might also provide the traction required to launch a successful identity metasystem.

2.8.1 Facebook as an IdP

Facebook[13] is a hugely popular social networking site with over 500 million active profiles [20], it stores a considerable amount of personal data about users (name, address, email address, instant messaging ID, phone number, and photographs), it also

[13] https://www.facebook.com/ (accessed 15 May 2011).

maps relationships with friends, and events the user has been invited to. Much of the personal data an IdP would require is already held by *Facebook*.

The *Facebook* platform is a set of Application Programming Interfaces (APIs) that allows third party developers to integrate into *Facebook* and access its core features. The platform components of particular interest to IdM are

- The Graph API – provides access to the 'social graph' - objects (users, photos, events, pages, etc) and the connections between them (friendships, shared interests ('Likes') shared content, photo tags).
- Authentication – Single-Sign On (SSO) across web, mobile and desktop applications.
- *Facebook* Connect – enables *Facebook* members to login to third party applications using their *Facebook* identity.

These components align closely with some of the features necessary for an identity metasystem; the Graph API holds the information and allows legitimate relying parties access to it. The authentication mechanisms provided a SSO in a unified customer experience, and *Facebook* Connect allow the identity to be federated.

When a user installs an application into their profile the application lists the permissions (the access to other data) it requires and the user is able to withdraw from the transaction at any stage. SSO is naturally part of any identity management system, but *Facebook* Connect enables users to authenticate to thirds parties (relying parties) without a separate registration process. The login credentials remain inside *Facebook*, the third party in merely assured the user has been adequately verified by *Facbook*, this approach helps increase security and privacy for the user.

Facebook is in reality lacking some key IdP qualities, namely security and privacy. It has leaked data through misconfigured systems [21], has allowed developers easy access to unnecessary personal information [22] and plays host to SPAM and other social engineering attacks [23].

At the moment and, in the author's opinion in the short to mid-term *Facebook* is not suitable as an IdP for financial transactions and it is difficult to imagine a bank entering into a partnership with them. However it is likely to see increased use as a low value identity broker. If *Facebook* tightens or adapts it security model (perhaps firewalling off certain details) this could see it become an IdP *Facebook* would have difficulty adapting its business model to increase privacy (indeed its business model is uncertain anyway [24] [25]), so if it were to become an IdP it might not be a 'model' example as envisaged and described.

2.8.2 Government Intervention

The late-2000s saw a series of banking crisis throughout the world, with banks being propped up by governments using a range of economic measures. A similar crisis (probably with a large Internet dimension e.g. a loose of confidence in banks as a result of massive Internet fraud) could see governments stepping in again but requiring the banks to support an identity metasystem (perhaps by becoming IdPs as part of the bail out).

2.9 Business Impact

Federate identity management has made significant inroads in the enterprise and academic environments; here the organisation provides and controls the user's identity and typically serves as the IdP. The organisation funds these services as they provide superior security and efficiency. The organisation also controls the infrastructure where the identity is used or has contractual agreements with partners when the identity is used beyond its immediate control. This differs from the consumer model; consumers are unwilling to pay for a technology they perceive as currently being free, and relying parties are unwilling to surrender their identity silos which are a rich source of sales and marketing information. While the technology exists and is suitably mature, without the emergence of a sustainable business model the status quo will continue.

This work provided BT with a clear understanding of how the consumer market differs from that of the enterprise and academia, and how without a catastrophic breakdown in Internet trust and security this market cannot be readily capitalised. BT should remain current with the technology and continue to observe developments

(technical, legal and social), but not launch any products or services until a viable revenue stream emerges.

2.10 Identity Paper

The following paper, in addition to being reported within BT was presented at *The Fourth International Conference on Legal, Security and Privacy Issues in IT Law* (LSPI) and published in the *International Journal of Liability and Scientific Enquiry*. It expands on some of the ideas presented in this chapter.

Future Vision of Identity

T. Martin, C. Durbin, M. Pawlewski, and D. Parish
T. Martin, C. Durbin, M. Pawlewski
BT, e-mail: {thomas.2.martin, chris.durbin, mark.pawlewski}@bt.com
D. Parish
Loughborough University, e-mail: d.j.parish@lboro.ac.uk

Abstract This paper presents two possible visions of the future, one where the status quo of consumer Identity Management remains largely unaffected by the emerging technologies, the other where these new technologies gain popularity and wide-spread acceptance. More importantly, we give arguments that support both futures occurring. The object of this exercise is to try to enumerate those obstacles that lie in wait for the mainstream acceptance of any significantly new approach to Identity Management, before we can determine how they are overcome.

1. Introduction

There is a commonly held belief that the management of digital identity online will be 'the next big thing'. There is no shortage of literature on the subject, nor supporters for the cause[14]. It is generally agreed that a new approach to Identity Management (IdM) in the digital world would be beneficial (in this paper, we refer to IdM as the management of individual's identity as they join, leave and interact with different organisations). The enterprise space has many Single Sign-On (SSO) federation tools such as PingFederate[15]
and Active Directory Federation Services[16]; Shibboleth[17] has made inroads in academia; consumer IdM services include CardSpace[18], OpenID[19], Sxipper[20], and Higgins[21]. Despite the availability of these technologies to assist in implementing new IdM strategies, there is no clear evidence that these approaches to IdM would be adopted on a wide scale basis in the foreseeable future (Bennett, R. 2009).

There is overwhelming evidence that current IdM is failing us. Many internet websites now require some form of registration of user credentials and individual users are now faced with the task of remembering a number of usernames and passwords. This multiplicity of user accounts has exploded to an extent where user frustration known as 'password fatigue' frequently leads to insecure practices such as employing the same username and password across multiple sites i.e. different websites and services. At the root of the problem is the fundamental flaw that the internet was not designed, but evolved without a uniform system of digital identity in place. There have been numerous attempts to solve this problem, such as Microsoft Passport, but many of these have failed leaving a scattering of inconsistent, ad hoc, partial solutions (Bertocci, V., Serack, G., & Baker, C. 2008).

However, the promise of the new crop of Identity Management systems extends beyond managing the glut of usernames and passwords. One of the wider challenges is to provide users with immediate

[14] See http://identityblog.com (accessed 07-06-08) or http://confusedofcalcutta.com (accessed 09-07-08).
[15] http://www.pingidentity.com/products/siteminder.cfm (accessed 12-07-08) .
[16] http://www.microsoft.com/windowsserver2003/r2/identity_management/adfswhitepaper.mspx (accessed 14-07-08).
[17] http://shibboleth.internet2.edu/ (accessed 15-08-08).
[18] http://www.microsoft.com/net/windowscardspace.aspx (accessed 11-08-08).
[19] http://openid.net (accessed 19-06-08).
[20] http://www.sxip.com/sxipper (accessed 23-08-08).
[21] http://www.eclipse.org/higgins/ (accessed 12-06-08).

access to websites sites where they have not already enrolled but do meet the requirements for access, e.g. being over 18 years of age and possessing a valid credit card. In this type of scenario a user would need to be enrolled with an Identity Provider (IdP) who performs the task of mediating between the user and
the web site (Relying Party (RP)). Although there is clearly a need for new Identity Management systems that meet the challenges of the modern internet, it is evident that users are not ready to use such systems (Harrington, E. 2004). There are currently only a handful of IdPs available on the internet and these provide very limited functionality. There is no sign, now, nor on the horizon, that an IdP will emerge and provide genuinely useful functionality such as access to e-commerce sites (e.g. amazon.com), or online banking.

There are many reasons why the small IdPs that do exist are currently confined to activities such as blogging communities and social networking sites. We explore both, these reasons, and also the forces pushing for advances in IdM.

The aim of this paper is to explain why early attempts at IdM have failed to find grassroots acceptance, why the current status quo of separate information silos is unsustainable and what road blocks exists to a successful Internet-wide IdM landscape.

1.1. Document Structure

Section 2 gives an overview of the state of the art in Identity Management. Section 3 splits into two parallel threads. The two threads represent two contrasting possibilities for IdM. Section 4 then provides guidance on how progress can be made, based on the arguments and possible consequences from Section 3.

2. Identity Management Overview

In this paper, we look at a number of different approaches to the management of identities and the developing trends they are taking. This covers the complete life-cycle of digital identities, as well as related issues of federation, privacy and standards. While we discuss different technologies, we are largely concerned with higher level issues. The main differences we wish to explore are between current, well established methods, and newer experimental approaches that are being put forward. Current methods largely use username/password authentication tied to an email address, coupled with detailed personal information on the user, and restricted to a single domain (siloed). The new Identity Management approach/technology/framework will be discussed in Section 2.1, but broadly attempts to give more control to the user and enable better cooperation between organisations. The IdP manages the identity for the user and can attest their identity to the RP. The RP is spared the trouble of user administration, and the user's convenience can give a better customer experience.

From the technology perspective, interactions can be viewed as being enterprise or consumer. In this context, enterprise interactions refer to those where the general pubic is not involved. As such, these interactions would be between companies, governments or between companies and governments. Consumer interactions, on the other hand, refer to interactions where at least one of the parties is an individual from general public. The distinction between consumer and enterprise is made here because it has a bearing on the type of technology that can be deployed. Enterprise Identity Management is beyond the scope of this paper.

2.1. Consumer Identity Management

The situation for consumer applications (consumer as defined above) is much more complicated in that the scope of the problem is less well defined; interactions are more complex and the end user, i.e. the general public, has got to 'buy in' to the proposition. As such there are very few advanced Identity Management systems currently available, and the technology is still in its infancy. There are many different approaches to this problem, but the consensus favours the user-centric model (Cameron, K. 2005). User-centric Identity Management is based around the premise that identity is personal information and should be under the control of the identity subject. With the current model, Identity information is scattered across a wide number of organisations with the user having little control over how this information is used.

2.1.1. Technologies of Note

Two technologies feature heavily in this paper - Microsoft CardSpace (Chappell, D. 2007) and OpenID. CardSpace is a cryptographically hardened approach based on extensible open standards, with support for third party IdPs and user generated self-issued cards. OpenID is a lightweight, browser driven, SSO tool based on existing web protocols. Inherently insecure ('phish-able'), it has nonetheless found favour in the blogosphere. OpenID now features CardSpace support - maintaining the ease of use whilst being hardening against attack[22] (Fox, K. 2007) (Jones, M. 2008) (Recordon, D. 2007).

Higgins[23] is an open source project from the Eclipse Foundation which presents an umbrella approach - a flexible data model, portable across platforms, inclusive of multiple technologies and protocols. Higgins aims to unite various approaches and deliver an identity meta-system.

2.1.2. Main features for successful IdM

Federation – A user should be able to enrol with a provider and use their identity anywhere, much as they would with a credit card. A credit card is provided by a financial institution but can be used anywhere.

Login Security – Passwords can be a weak link in the login process, multi-factor authentication is a good method for mitigating this weakness. Hardware devices such as One Time Password (OTP) 'key fobs', or transmitting an OTP to a mobile phones are often used for this.

Information Security – Loss of customer records leading to identity theft is a source of concern to consumers. They must have confidence in their provider, this confidence stems from good business practises backed up with secure technology.

User Centricity – Users want to be in control of their information, they should be aware of what information RPs are requesting, and should be able to cancel a transaction if they feel the RP is being too invasive.

2.2. Common Goals

Looking at current IdM solutions, it is clear that there are a number of common goals. First is some sort of *Federated Infrastructure*. Members of the public access numerous domains on a daily basis. Federation at the least ensures a consistent experience, and can provide Single Sign-On (SSO). Users enter their authentication credential's once, to gain access to a range of sites.

Most solutions are also promoting the *user-centric model*. This puts the user at the centre and in control of all transactions that involve any aspect of their identity. All transactions from the IdP to a RP are approved by the user before being passed onto the RP, and the reverse is also true. There may also be a trust relationship between the IdP and the RP, but this is only essential for transactions where the RP requests specific information that
requires some form of underwritten guarantee.

Interoperability between technologies is an essential ingredient for the widespread take-up of IdM. If IdM is to become pervasive it is likely to stem from disparate technologies being able to interact via a common infrastructure, and able to cope with different technologies and users having multiple identities. Within a consumer context users may want to support a range of identities, for instance purchasing of adult content, online gaming, etc, may be transacted under a different alias to your employment identity for instance. As individuals conduct more business and leisure online the ability to separate these personae and maintain privacy will become increasing important (Fish, G. 2009) (Bennett, R. 2009). Work in this area includes WS-*[24] and Liberty Alliance.

3. The Prospects for Identity Management

[22] http://www.identityblog.com/?p=668 (accessed 10-10-08) and http://www.identityblog.com/?p=659 (accessed 10-10-08).
[23] http://wiki.eclipse.org/Higgins (accessed 25-09-08).
[24] http://www.w3.org/2002/ws/ (accessed 30-09-08).

This section primarily concentrates on Consumer IdM, which has yet to realise its potential. The sections present two possible directions for Consumer IdM.

Figures from the United States Federal Trade Commission (FTC) show cases of reported identity theft increased from 31,140 case per year in 2000, to 313,982 in 2008, see Figure 1. Although there is much variation in the figures during that period the overall trend shows identity theft increasing. The real number is likely to be considerably greater, the figures only represent cases reported to the FTC - people may simply report suspected cases of ID theft and fraud to their bank and credit referencing agency. The FTC data (Finklea, K. 2009) supports an argument that the current approach to IdM is failing, and will continue to see identity fraud increasing until it plateaus out and remains at a level that fails to provide the critical mass necessary to drive the adoption of a new identity meta-system. Alternatively identity fraud may become so endemic that the public will embrace a more radical approach, even if this leads to even more fraud in the short-to-medium term as the system is rolled out.

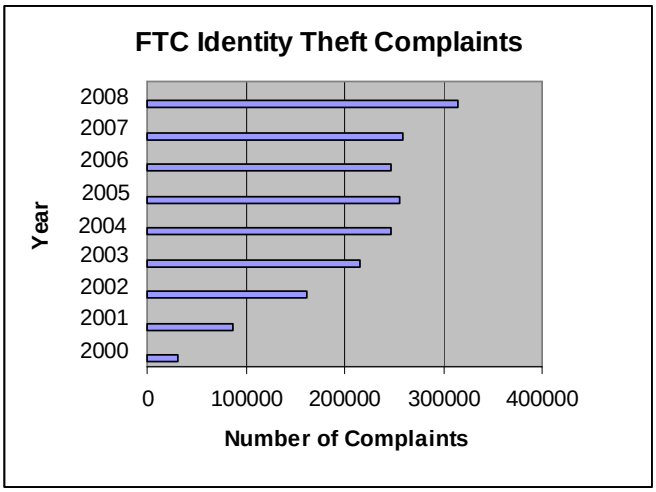


Figure 1 – FTC Identity Theft Complaints

3.1. Bleak Future

The future predicted in this section is one where all the new technologies being introduced in Identity Management fail to take off. The existing siloed approach of single domain identities remains in place. The main argument given for this is user acceptance. The approach taken here is to try to learn from past experiences about what has been done to change user behaviour and to apply this to Identity Management.

3.1.1. Internet Revolutions

In its short history, the Internet has gone through many changes. Some of these changes have been slow and gradual, but most have been new technologies (or uses of existing technologies) that have become extremely popular, very rapidly. To capture the latter type, we define the word “revolution” to mean any technology that:

- Changed how people used the internet
- Was very different from what came before
- Became very popular, very quickly

The emphasis is on user experience. The technology itself need not be revolutionary, nor is it limited to looking at commercially motivated/driven revolutions. Nevertheless, technological advances are typically components of these revolutions and new business models quickly spring up around them.

3.1.2. Properties of a successful Revolution

The following are a list of technologies that qualify as revolutionary, (as defined above) - www, e-mail, search engines, VoIP, social networks and MMOs/Virtual Worlds.

The choice of these seven technologies as revolutionary is, we hope, uncontroversial. Each has changed the way people use the internet (although strictly speaking, the web and e-mail did not change how people used the internet so much as cause people to use it). Each is significantly new in terms of how it is used (real-world parallels will be discussed later), although some may be arguable (e.g. VoIP has similar predecessors in Instant Messaging, IRC, etc., but the speed, convenience and usability differentiates it). The time it took each of these technologies to become popular has certainly varied. But each of them has enjoyed a period of rapid growth to the point of receiving media attention. Rather than failing to live up to the hype, each has continued to grow in popularity. There are several characteristics that are clearly common across all these revolutions:

Each of the technologies listed has an *equivalent real-world activity*. E-mail equates to mail, VoIP to telephone, search engines to directory enquiries, social networks on-line are just another type of social network. These parallels with the real world may increase take-up by making it easier to understand. This does not contradict the requirement that each revolution be very different to what came before.

Naturally, these technologies are not identical to their real-world counterparts. It is not enough that each differs, but in order to become popular they all had to *provide advantages over their real world parallel*. The main improvement is speed (consider the difference in time it takes to deliver email versus regular mail), but there can also be increased convenience (search engine) and reduced cost (VoIP). The technology must also provide a functional benefit and not be a technological fad. Immediate acceptance is not necessary but people should be able to envisage themselves using it.

Each of the technologies listed serves a *desired purpose*, which could be considered a consequence of the parallel in the real world. It must be something that people want and will use, not just technology for the sake of technology.

3.1.3. Why IdM will not be revolutionary

Most of the criteria for the definition of an 'Internet Revolution' can only be judged in hindsight. When applied to IdM, how much it will change how people use the internet will need to be seen, and how quickly it grows in popularity is what we are trying to determine. The third criterion, that it must be very different to what has gone before, is open to debate. The solutions being proposed in the field of Identity Management are all trying to replace the existing widespread use of individual username/password. No matter how different the new user experience will be, at heart they will still be "logging in" to a service. This already suggests that IdM will not be revolutionary, at least as far as our definition goes. The broader significance of this is that there is an existing solution that for all its flaws "does the job". Replacing this solution will require changing the habits of those who are of the mindset "if it's not broken, don't fix it". Having described in the previous section those properties that we see as common to all Internet Revolutions, the following discussion justifies how they largely do not apply to IdM.

The problems the new identity frameworks are addressing (weak authentication, proliferation of personal data, masquerading and misdirection attacks) exist in some form in the real world, but are nowhere near as serious or as widespread. None of the proposed solutions can be said to have true real world counterparts. One may argue that CardSpace, with its credit card like interface for the card selector, is an embodiment of a wallet. Authentication in CardSpace is not based on a physical object that can be used
anywhere, and it is strongly tied to the host computer. Using your CardSpace 'wallet' on any computer (as you would use your actual wallet in any store) is being researched, but it has not been at the core of the original design (Bertocci, V., Serack, G., & Baker, C. (2008). The addition of portability will certainly introduce problems in usability and/or security.

One approach to identity management that would make everyone's life easier is universal SSO from a single provider. All users would enjoy a consistent experience, would only need to authenticate once, and could be protected from attackers by a trusted entity. This is the Microsoft Passport approach. All subsequent proposals have pointed to its failure and worked on how to avoid repeating it. There will be a plurality of Providers, and a plurality of technologies that interoperate. Each of the proposed systems would no doubt curtail the password fatigue people now experience, but forcing users to learn and use many new technologies in parallel will undo this benefit. The CardSpace interface, for example, is visually very different from a website username / password. Even OpenID, based entirely on web protocols refers the users back to the OpenID provider. This maybe daunting and confusing to the nave

user. There is also an inherent problem with complexity. The definition of user-centric places the user in control. On the one hand, this can increase security, provide privacy, anonymity, control of acceptable use/transfer of personal information. On the other hand, being placed in control, the user now had to make decisions on these subjects. This control does have benefits, but convenience is not one of them.

The sales pitch for any of the new identity frameworks is a hard one. One selling point is security, but that has always struggled to get the population at large excited. Convenience is arguable, depending on how many solutions gain acceptance. The question is, how do you convince the man-in-the-street to adopt any of these new technologies? They serve a purpose, but because it is a purpose that is already being served (admittedly in a less than satisfactory manner) there is a momentum that needs to be fought.

3.1.4. What this means for IdM?

We have described many properties of several internet revolutions and given justifications of how these have contributed to their success. We have argued that these same properties are not shared by the new Identity Management approaches. While not being sufficient to claim that these approaches are doomed to failure, we believe we can state that they will not enjoy anything like the rapid growth in popularity of other technologies. These are very real hurdles to gaining widespread acceptance and either they must be mitigated, or a more aggressive deployment must be used.

3.1.5. Business Model

There are no IdPs available for sites where financial transactions take place and there are certainly no IdPs for internet banking applications. It is abundantly clear that the public would not pay for an IdP service and RPs such as retailers or banks are also unlikely to pay to provide such a service for their customers. Customer records provide a rich source of marketing information, organisations will be reluctant to surrender control of this to an IdP[25]. Aside from the fact that no financial mechanism to support the growth of an IdP has been found so far, there is also the issue of liability if the IdP data becomes compromised. An IdP would have to underwrite the risk and potentially have to pay significant compensation to users if things go wrong. In a nutshell the case for becoming an IdP is to provide an IdP service that no one is currently willing to pay for and underwrite the risk if things go wrong. The current situation with banks and credit card companies is that many of them absorb the losses for their customers who are the victims of banking and credit card fraud. From the customer's point of view this arrangement works very well. If an individual is subjected to fraud, then at most they will suffer the inconvenience of re-establishing their bank or credit cards, but are unlikely to suffer financially. There is therefore very little incentive for a banking customer to use an IdP (if one existed), unless it was free and also offered at least the current level of protection that financial institutions offer. This lack of business model is the crux of the problem and is arguably the main reason that there is currently no serious internet IdPs.

3.1.6. Bleak future summary

If IdM had many properties in common with the previous revolutions we discussed, then its success would seem likely. The fact that it does not is no proof that it will fail. But it does raise a valid question: "How can we convince the population at large to adopt a significantly new technology requiring a similar level of effort as they have been asked before, but without as clear a benefit?"

3.2. Promising Future

In this section, we describe how we see a new identity framework being established. We will not be trying to directly counter the arguments of the previous section, but rather assume a watered-down version of their conclusions. The new identity framework will not arrive quickly on a wave of popular support, but there are forces pushing us in this direction. We will describe these forces and show how they will eventually reach the common goals identified in Section 2.2.

We start this section with some very high level arguments, describing various trends that support our case. We also delve into specific areas that are moving us towards new approaches in Identity

[25] Google checkout (http://checkout.google.com accessed 06-01-09), allows customers to buy products from retailers. This is more of a payment service than an IdP and its success is tied to both the Google brand and the incentives Google provide for participating sellers.

Management. Finally, we look at the success of OpenID, and how it already provides many of the desired features of the new identity framework.

3.2.1. The Identity Management J-Curve

Section 3.1 described many hurdles to the successful adoption of an identity framework. We will describe how a major catalyst that could cause significant change in Identity Management, namely in the form of some catastrophic breakdown in security on the Internet. Although millions of people have had their personal information exposed either by identity theft or poor information management (Oates, J. 2007) and security practises the authors do not considered this a catastrophic breakdown. Very few people (as a percentage of the number of lost records) have incurred any financial loses or suffered any detriment to their credit history. Certainly not in sufficiently high numbers to overcome the inertia and apathy currently surrounding IdM

In order to understand the conditions that need to exist for widespread take up of new IdM technology, it is important to understand the J-Curve phenomenon. This basically predicts that things have to get worse before they get better and can be applied to a wide range of areas such as economics, medicine and political stability. In his book "The J Curve - A New Way to Understand Why Nations Rise and Fall" (Bremmer, I. 2006), Ian Bremmer explains that countries follow a J-Curve when moving from a state of isolation to a state of openness. Similarly, the J-Curve can be proposed as a way of modelling the potential changes in Identity Management.

A J-Curve for Identity Management is shown in Figure 2. In this context, the Openness axis refers to the degree of openness of the technology (standards vs. proprietary) as well as of the approach (use of Federation, transparency, etc). Moving to the right along this axis implies fewer standalone siloed approaches to identity and moves towards federated approaches. The Stability axis refers to the degree of security of the user experience, as well as the durability of the parties involved. Movement up the y-axis implies increased identity security and decreased volatility.

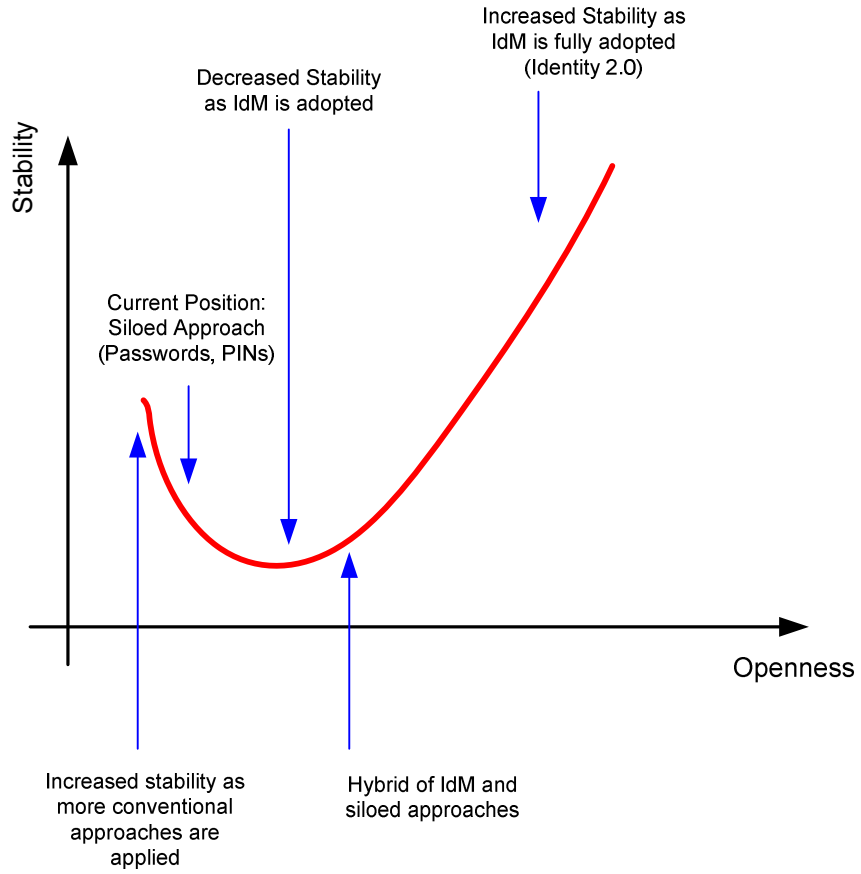


Figure 2: J-Curve representing IdM adoption

The current position of IdM is somewhere to the left of the lowest point on the curve. This is an interesting position to be at, because it can be seen that immediate increased stability would only be obtained by moving to the left. This implies the application of additional siloed security measures. Interestingly, this is precisely what is happening with current banking security. The introduction of chip and PIN is an example of this, and more recently, in the UK, the introduction of EMV (Europay, Visa, MasterCard) card readers and one-time password tokens for secure banking internet access. Nevertheless, with the introduction of increased siloed security measures comes an increased "user fatigue", this time not by the need to remember a multitude of passwords, but by the need to remember to carry a variety of hardware paraphernalia to facilitate secure internet access.

In the long term, it would be more desirable to move to the right. An open platform can give a simpler user experience, provide transparency and minimal disclosure (reducing the problems of identity abuse), and place the user in control. But opening up the technology will, in the short term, cause problems experienced by any new technology. In the early change-over period, users will inevitably make mistakes and become confused with the new technology. At this stage the technology is relatively immature and will certainly contain vulnerabilities. This can be seen in the OpenID standard, which is well known to be susceptible to phishing attacks. This is the dip in the J-curve, and the reason why the long right side of the dip is a longer slope: it is more difficult to take this path, but ultimately more rewarding.

3.2.2. Gradual Move Towards New IdM Technologies

It is estimated that a quarter of UK adults have had their identity stolen or know someone who has fallen victim to ID fraud[26]. With these figures increasing yearly it is entirely possible that Internet users will, perhaps reluctantly, start to use new IdM systems in order to protect their electronic interests. This take up would be more analogous to the take up of anti-virus software or anti-spam systems, being necessity rather than popularity driven. No one can say for sure what level Internet fraud would have to reach before widespread adoption, but it would seem that current levels are insufficient.

The entrenched usage of username/password is simply not scalable, and not sufficiently secure. But once users start to "vote with their wallets" against password fatigue and identity theft, there may be an impetus to change. This pushback by the users to bring about positive change, as we describe in the next section.

3.2.3. User Education

A definite barrier to the take up of new IdM technologies is the level of awareness of the public. This is such a problem that there are now government backed campaigns designed to raise awareness of the general problem of internet fraud and identity theft. Get Safe Online is one such campaign that provides straightforward advice for the general public[27]. The level of this advice is extremely basic, advising people not to publish their identity information (name, phone, number, home address) on public web pages or social networks such as Facebook. The fact that it seems necessary to advise the public on some very fundamental steps to protect their identity perhaps highlights the scale of the task. On one hand there are the IdM experts talking about the merits of IdPs, SSO, OpenID and CardSpace technologies and on the other is the general public leaving their personal information on websites for the world to see. There is a chasm between the two groups and this clearly needs to be bridged before the general public will accept and use new IdM technologies.

3.2.4. Compliance with Legislation

With internet-based identity fraud growing rapidly, it is feasible that the Government could mandate compulsory IdM technologies for online financial transactions. The exact makeup of these technologies is open to debate, but this is already happening to an extent with the introduction of the APACS (Associates for Payment Clearing Services) specification for two-factor online authentication standard which stipulates that banks should use two factor authentication for banking transactions.

[26] http://news.bbc.co.uk/1/hi/business/4311693.stm (accessed 24-03-09).
[27] http://www.getsafeonline.org (accessed 02-08-08).

3.2.5. Learning from OpenID

OpenID is by far the most popular IdM system available on the Internet. The most important positive aspect of OpenID is the fact that it is not a centralised identity service, users can choose which IdP they are happy to entrust with their identity information. This decentralised nature means that the whole system does not fall apart if one IdP goes out of business. Users can even set themselves up as their own IdP. The main aspect of OpenID that attracts users is in its ease of use, with only one user name and password to remember.

There are also positive benefits to the website operators in that they can simplify user registration, a well known obstacle for signing up new users. Organisations are protected in that they no longer need to store personal identity information which potentially could be lost or stolen from their server (causing both financial and reputation damage).

OpenID has a strong following and a very supportive community. While it is generally agreed that OpenID does not address the real problems in Identity Management it is still relatively young and there is time for the OpenID technology to evolve into a credible IdM technology, and for others to learn from its success.

3.2.6. Promising future summary

There is no doubt that the battle for the general acceptance of new IdM technologies is only just beginning. The evolution of these is potentially going to stem from the current trivial uses such as OpenID in social networking and in blogging applications. Popularity of the technologies in these communities could fuel a demand for widespread use of more advanced, secure systems. The fact that OpenID is notoriously insecure is somewhat irrelevant, people are using it. The desire for ease of use and convenience will
fuel demand for new IdM technologies. Systems based on OpenID and Microsoft CardSpace will no doubt emerge as the front runners and with this, the technology will evolve to meet the stringent security requirements required for commercial/financial transactions.

4. Conclusion

We have discussed two potential, contrasting, futures for the management of identity in the electronic world, and have given rational arguments to support both views.

The Bleak Future predicts that very little will change in the foreseeable future and asks the question: “How can we convince the population at large to adopt a significantly new technology requiring a similar level of effort as they have been asked before, but without a clear benefit?” The Bleak Future, although perfectly feasible and valid as a potential outcome, is not very interesting from the research perspective. If it is believed that this outcome is likely, then it would seem that the best course of action would be to resist wholesale overhauls of identity infrastructures, and make minor, iterative changes to existing systems.

The Promising Future predicts that new identity management techniques will slowly start to grab a foothold and will gain widespread acceptance throughout the digital world. If this direction does prove to be the correct one, then this raises a number of interesting questions both with regard to technology and to the emerging attitudes towards digital identity. The predominant players (Card Space, Liberty Alliance and OpenID), may find opportunity to combine their technologies into an all encompassing meta-system. Any security shortcomings in these systems could be shored up with the relevant technologies, i.e. biometrics, hardware tokens, etc.

Despite the pessimistic outlook of Section 3.1, there is no denying that there is a need for a change to the way we manage our personal identities on-line. By being wary of the pitfalls and possible road-blocks, a safer and more prosperous future can be made real.

In the short term the identity ecosystem will follow the path of the bleak future, seeing a proliferation of username/password, OTP and other proprietary siloed approaches, pitting service providers against thieves and fraudsters in a game of cat-and-mouse. If service providers can keep fraud to an ‘acceptable’ level the status quo may remain in place. If identity fraud increases to levels that cannot be tolerated it is likely that a new identity meta-system as described in the promising future will emerge.

References

1. Bennett, R. (2009). *Plea to ban employers trawling facebook - times online.* Retrieved 10/02/2009, from http://technology.timesonline.co.uk/tol/news/tech_and_web/article3613896.ece

2. Bertocci, V., Serack, G., & Baker, C. (2008). *Understanding windows CardSpace: An introduction to the concepts and challenges of digital identities (independent technology guides)* Addison-Wesley Professional. Retrieved from http://isbndb.com/d/book/understanding_windows_cardspace_an_introduction_to_the_conce

3. Bremmer, I. (2006). *The J Curve: A New Way to Understand Why Nations Rise and Fall*. New York: Simon & Schuster.

4. Cameron, K. (2005). *The laws of identity.* Retrieved 21 May 2007, from http://www.identityblog.com/stories/2004/12/09/thelaws.html

5. Chappell, D. (2007). *Introducing windows CardSpace.* Retrieved 15/08/2007, from http://msdn2.microsoft.com/en-us/library/aa480189.aspx

6. Fish, G. (2009) *BusinessWeek debate room employers, get outta my facebook.* Retrieved 10/02/2009, from http://www.businessweek.com/debateroom/archives/2008/03/employers_get_o.html

7. Fox, K. (2007). *Janrain blog: MyOpenID adds information card support.* Retrieved 22/10/2008, 2008, from http://blog.janrain.com/2007/10/myopenid-adds-information-card-support.html

8. Finklea, K. (2009). *Identity theft: Trends and issues.* Congressional Research Service.

9. Harrington, E. (2004). Identity Management in the Open Group. *SEWP Identity Management Symposium*

10. Jones, M. (2008). *Mike jones: Self-issued » re: OpenID kills windows CardSpace?!* Retrieved 22/10/2008, 2008, from http://self-issued.info/?p=59

11. Noguchi, Y. (2006). *Access denied - washingtonpost.com.* Retrieved 09/02/2009, from http://www.washingtonpost.com/wp-dyn/content/article/2006/09/22/AR2006092201612_pf.html

12. Oates, J. (2007). *Darling admits revenue loss of 25 million personal records • the register.* Retrieved 20/02/2009, from http://www.theregister.co.uk/2007/11/20/hmrc_loses_lots_data/

13. Recordon, D. (2007). *OpenID » blog archive » microsoft and google both ship OpenID.* Retrieved 22/10/2008, 2008, from http://openid.net/2007/12/03/microsoft-and-google-both-ship-openid/

14. ZDNet.co.uk. (2005). *Passport failure shows the folly of microsoft's ways - ZDNet.co.uk.* Retrieved 06/02/2009, from http://news.zdnet.co.uk/leader/0,1000002982,39183062,00.htm

3. TEXT ANALYSIS

More and more children are using the Internet for both education and leisure. Many children converse with friends in chat rooms, by Instant Messaging (IM) and via social networks. This may entail the continuation of 'real world' friendships online or meeting contacts and forming relationships in a purely Internet setting.

In the real world it is relatively easy for parents to meet and keep track of their children's friends. On the Internet this is more difficult. People can easily conceal their true identity and adopt a completely different persona. Groomers and abusers who wish to harm children are known to communicate with them via text based mediums. It is important to ensure children are protected from the dangers that such individuals pose.

The following chapter investigates the feasibility of developing algorithms that can differentiate between children chatting to their genuine friends and the same children chatting to adults attempting to groom them. The basic premise here is that the style of text chat between genuine friends is likely to be different to that of text chat between a child and a groomer. The extent of this difference will certainly vary from case to case. At one extreme the groomer can be very blatant and direct using a language style that can reasonably easily be detected via simple word/phrase spotting techniques. Statistics from [1] suggest that 80% of groomers fall into this category. At the other extreme the child groomer can be much more subtle in their approach, it is this type of grooming that is of interest here. In considering subtle child grooming, the main questions are 1) to what extent can subtle child grooming be detected using an automatic system? and 2) can existing techniques be improved?

The general problem of detecting grooming activity in text chat is similar to that of detecting spam email. In both cases (spam detection and grooming detection) the challenge is to detect text that doesn't fit the model of a bona fide email, for the case of spam, or bona fide text chat for the case of a child groomer. Arguably, the degree of difficulty of detecting spam is similar to that of detecting a reasonably blatant child groomer, but as the grooming style becomes more subtle, the difficulty of detection increases. Spam detection is also easier in that generally for a spam email the majority

of its content will indicate that it is spam, but for grooming text chat, only a small fraction is likely to be blatant sexual grooming.

3.1 Text based Internet conversations

Message boards, news groups, email, IM, chat rooms and social networks allow users to participate in conversations on the Internet. These conversations can be synchronous, where the text chat is a two-way real time exchange, or they can be asynchronous where the chat is over a message board where several hours (or longer) may elapse between entries.

Instant Messaging, a synchronous chat tool, has experienced an explosive growth in recent years. In February 2006, 82 million Europeans (49% of the population) used IM [2]. In Canada 56% of children use IM, 27% use it every day [3]. 58% of children aged 12-17 have a social networking profile [1].

Modern media rich Internet applications are blurring the traditional lines between IM, chat rooms and social networking sites. The key issue is that all use text based communications. Statistics [1] indicate that 77% - 86% of online solicitations occur over chat or IM sessions.

3.2 Automatic Grooming Detection

The following section describes a set of experiments that were designed and run in order to establish the feasibility of detecting grooming through text based channels. The immediate difficulty in designing such tests is that there is not a sufficient text corpus available to evaluate a) build models to represent grooming style text chat and b) testing the robustness of these models.

There is one well known test corpus, the Perverted Justice (PJ) corpus that is publicly available but the text in this corpus falls into the category of being highly blatant text chat exchanges which would be easy to detect with simple word spotting techniques.

In order to create a test which is more realistic than the PJ corpus, a number of texts from different authors were chosen. The task was to establish if it is possible to

differentiate between the writing styles of different authors. The rationale being that this task would be the same level of difficulty as that of differentiating between the writing style of a child and that of a groomer masquerading as a child.

Four source texts were selected - *On the Origin of Species - Charles Darwin, Fahrenheit451 – Ray Bradbury, Nineteen Eighty-Four - George Orwell* and *The War of the Worlds – HG Wells*. Three of these texts *Fahrenheit 451, Nineteen Eighty-Four* and *The War of the Worlds* are broadly similar works of science-fiction. *On the Origin of Species* is substantially different – it is a piece of scientific literature that introduced the concept of biological evolution.

This makeup was chosen to determine if the algorithms would detect that one text was substantially different from the others. The assumption here is that harmful chat would be substantially different to children's everyday chit-chat. *It is of course not ethical to expose children to harmful material in order to test classifier performance*.

The texts are of various lengths but were normalised to the length of the shortest work (*Fahrenheit 451).*

Each book was divided into multiple 500 word test sets and a single (per book) 5,000 word training set. Each book was divided in eighty-two[28] sequential 500 word test sets and one 5,000 word training set. In all books the training set began at word 15,000. No test sets where subsets of the training set [$testset \notin training$].

3.3 Classification Tests

Three categories of test were performed. These were:

- Simple trigram
- Bayesian
- Gaussian Mixture Models (GMM) using a novel feature extraction technique

Each is explained below.

[28] As a sanity check each training set was also tested again itself.

3.3.1 Simple trigram

As a baseline, a simple trigram[29] counter was tested. This examines each of the 498 trigrams in a 500 word test set and counts how many times each trigram appears in each reference training set of 5000 words. Thus if, for example, the first trigram is most popular in the work: *"On the Origin of Species - Charles Darwin"*, then this author gets a vote. This is performed for all trigrams and the author with the most votes is deemed to be the author of the unknown (test) text. Trigrams are used because they have been shown to be optimal in text classification [4]. Figure 5 illustrates an example of trigram harvesting.

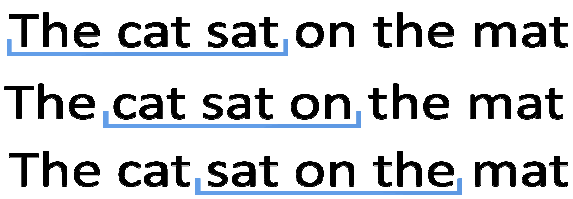


Figure 5 - Trigrams

3.3.2 Bayesian

Bayesian classification is considered by many to be the leading method for use in text classification (particularly in spam detection) [5] [6] [7] [8].

Bayesian analysis allows probabilities to be revised in the light of new evidence.

"*Bayes' theorem, in the context of spam, says that the probability that an email is spam, given that it has certain words in it, is equal to the probability of finding those certain words in spam email, times the probability that any email is spam, divided by the probability of finding those words in any email:*" [9]

$$p(spam \mid words) = \frac{p(words \mid spam)\, p(spam)}{p(words)}$$

[29] A trigram is a 'token' of three words.

Bayes' equation allows conditional probabilities to be calculated. Typically a classifier will measure the probability of a sequence of words being spam when matched against a spam model.

For the purposes of differentiation between authors, Bayes' equation is applied as follows:

There are four models *m1* to *m4* corresponding to the four authors and there are a number of test tokens (trigrams), *N*.

Each test token is compared against each of the four models. This gives four *P(o|m)* values *P(o|m1), P(o|m2), P(o|m3), P(o|m4).*

Strictly speaking *P(m|o)* is required, but in reality the actual value of *P(m|o)* is not required. It is only necessary to know which of the four [(*P(m1|o), P(m2|o), P(m3|o), P(m4|o*)] is the highest. This then corresponds to which of the models, m1, m2, m3, m4 is the most likely.

The relative ranking of *P(m|o)* can be inferred from the relative ranking of four *P(o|m)* values [*P(o|m1), P(o|m2), P(o|m3), P(o|m4)*]. From Bayes' equation *P(m|o)* can be calculated from *P(o|m)*:

$$P(m \mid o) = \frac{P(o \mid m)P(m)}{P(o)}$$

P(o) is the probability of an observation *o*, irrespective of which author it comes from.

P(o) is not known. However, in this case it is constant and in general it would be reasonable to assume that it is always constant or changing very slowly.

Similarly, for this test $P(m_1) = P(m_2) = P(m_3) = P(m_4)$,

and

$$P(m_1)/P(o) = P(m_2)/P(o) = P(m_3)/P(o) = P(m_4)/P(o) = K$$

Thus, in general $P(m|o) = P(o|m).K$

This means that the highest $P(o|m)$ will correspond to the highest $P(m|o)$.

Each training and test set is divided into tokens – three word trigrams. The probability of each unique token appearing in the text i.e. $P(o|m)$ is calculated as the number of times the token appeared divided by the total number of tokens in the (training) dataset.

$$P(o \mid m) = \frac{NumberOfTimesTokenAppears}{TotalNumberOfTokens}$$

The probabilities for $P(o|m)$ are very small and therefore Log probabilities are used in order to maintain precision.

3.3.3 Gaussian Mixture Model (GMM)

This approach to classification is based on GMM techniques which are widely used in many areas of pattern recognition and notably in speech technology [10] [11].

The author developed a novel feature technique extract to generate vectors for GMMs. The technique produces a sequence of vectors, each vector having a number of coefficients. The following example illustrates the feature extraction technique. In this example four coefficients are used per vector.

Consider a sentence extract such as *'The cat sat on the mat'* taken from a large body of text. In vector form this can be represented as:

$$\begin{pmatrix} Th \\ The \\ The\_ \\ The\_c \end{pmatrix}\begin{pmatrix} he \\ he\_ \\ he\_c \\ he\_ca \end{pmatrix}\begin{pmatrix} e\_ \\ e\_c \\ e\_ca \\ e\_cat \end{pmatrix}$$

Figure 6 - Feature Vectors (Underscores represent spaces)

The first coefficient refers to letter pairs (tokens lengths of 2 letters), the second coefficient refers to letter triples (token lengths of 3 letters)and so on.

These vectors (figure 6) are further processed into a normalised frequency of occurrence.

Suppose the letters *"Th"* occur 100 times in the body of the text, the *"Th"* in figure 6 is replaced by 100. Similarly other letter sequences are replaced by their frequency of occurrence. If *"he"* occurs 75 times then *"he"* is replaced by 75. Figure 7 shows numerical coefficients.

$$\begin{pmatrix} 100 \\ 80 \\ 69 \\ 54 \end{pmatrix}\begin{pmatrix} 75 \\ 68 \\ 62 \\ 58 \end{pmatrix}\begin{pmatrix} 94 \\ 79 \\ 67 \\ 41 \end{pmatrix}$$

Figure 7 - Vectors representing frequency of occurrence

Clearly *"The"* will occur less that *"Th"*.

The final stage of the feature technique is to normalise the coefficients. Each coefficient is divided by the total number of equivalent size tokens in the texts i.e. the first coefficient would be divided by the total number of letter pairs in the text sample. Similarly the second coefficient would be divided by the total number of letter triples.

In general, if there are *N* characters in the text sample, then the number of tokens is given by

$$M = N - L + 1$$

Where *M* is the number of tokens and *L* is the number of letters in a token.

The algorithm is case sensitive and includes punctuation to maximise entropy.

3.4 Results

During development and in initial testing the GMM method appeared to function well and results were promising. With a larger, statistically significant data the method proved unreliable.

Table 1 lists the results by classification technique and author (book). The unreliability of the GMM method is clear. It is further apparent that the Bayesian method offers superior classification to simple trigram matching.

	GMM	Bayesian	Simple
Bradbury	1.2%	81.2%	67.1%
Darwin	1.2%	100%	100%
Orwell	55.4%	91.5%	89%
Wells	84.3%	85.4%	84.1%
Average	35.5%	89.5%	85.1%

Table 1 - Classification results matrix

As Darwin is substantially different from the other works it is expected a good classifier would detect this.

3.5 Discussion

Despite successful application in other pattern matching fields and the initially promising results the GMM method can be considered a failure based on these results.

The method should perhaps not be completely discounted. Statistical researchers may find that changing the token structure to words, altering the token length, or varying the number of coefficients may improve the quality of classifier. Word stemming and

stop word removal may also improve accuracy. The classifier may prove beneficial in other areas of text analysis such as text mining. Varying in the attributes in the GMM (number of modes) may also improve classification.

Although the headline figure for the simple technique is similar to the Bayesian score the classifier was significantly poorer at detecting *Bradbury.*

The initial assumption that detecting grooming is similar to detecting spam is supported by these findings – Bayesian classification appears the best technique at present. *This assumes that grooming chat (like Darwin in the experiments) is significantly different from regular chat. Even differentiation of the similar texts was encouraging.*

3.6 Conclusion

Chatting online is a popular activity among children and parents are rightly concerned about their children's safety and whom they might be chatting. Monitoring and reviewing a child's online conversation can be onerous especially if the child spends long periods online or uses lingo and slang with which the parent is unfamiliar.

In an effort to improve the safety of children participating in text based conversations on the Internet a novel feature extraction method was developed and combined with a GMM classifier. The hope was this approach would prove superior to existing classifiers, more accurately determining whether a child was chatting to a groomer. The approach was based on techniques commonly used in voice biometrics and speech processing.

Bayesian classifiers are popular for 'undisciplined' text (informal text, as used in 'everyday' conversation that may feature poor spelling, grammar and syntax) and have proved very successful in spam detection.

The novel/GMM method performed proved to be a wildly unreliable classifier, significantly worse than the current leading approach – Bayesian classification.

Statistical researchers may find interest in the approach and might be able to significantly improve the classifier, but the author believes that Bayesian represents the best choice in this environment.

The value of automated protection methods when social factors are considered is critiqued in chapter 5.

3.7 Business Impact

This work has demonstrated that the Novel/GMM approach to text analysis though an interesting research exercise is ineffective and even with the benefit of further work is unlikely to improve on the current leading method, Bayesian trigram analysis.

The systems analysis of the problem space (chapter 5) questions the 'real world' performance of technologies attempting to detect grooming and sexual predation. Blatant grooming can certainly be detected but the subtler form (where the groomer shows attention, interest and affection in their victim) are identical in nature to teen dating. While the providers of anti-grooming technology keep experimental methodologies, training data and test data secret their results should be treated sceptically. This work has shown BT that such technology is of questionable value and should not be pursued further at this stage.

3.8 Text Analysis and Message Review Paper

The following paper, in addition to being reported within BT was presented at *The Fourth International Conference on Legal, Security and Privacy Issues in IT Law* (LSPI) and published in the *Journal of International Commercial Law and Technology.* It describes a method for grooming detection using automated analysis including the techniques discussed in this chapter to flag messages of concern for anonymous human review.

A Pseudonymous Peer-2-Peer Review System for Child Protection On-line

T. Martin, C. Durbin, M. Pawlewski, and D. Parish
T. Martin, C. Durbin, M. Pawlewski
BT, e-mail: {thomas.2.martin, chris.durbin, mark.pawlewski}@bt.com
D. Parish
Loughborough University, e-mail: d.j.parish@lboro.ac.uk

Abstract Children are using the internet more and more, and from a younger age. This is despite the commonly known dangers of predators. There is no policing of the internet, nor would it be possible to instigate. Parents are in the difficult position of trying to monitor and control their children's internet usage, when more often than not the children know the technology better than they do. This can lead to either ineffective measures, or measures that the children will deliberately circumvent for their own privacy. There are also technical issues that are far from trivial. The problem of distinguishing the dialogue of a child from a predator pretending to be a child is extremely difficult. This paper presents a solution which can accurately identify threats, while satisfying the apparently conflicting needs for safety of, and privacy for, the children.

1. Introduction

This paper looks at the problem of protecting children from on-line stalkers/predators. A recent survey of 1,500 children (aged 10-17) in the United States found that approximately 1 in 7 (13%) received unwanted sexual solicitations, and 34% communicated online with people they did not know in person (Wolak, Mitchell & Finkelhor, 2006). This often took the form of crude or vulgar comments in chat rooms - the victims were not bothered and handled the situation well. However some victims felt traumatised and some are targets of aggressive online solicitations (Mitchell, K., Finkelhor & Wolak. 2007). There is also a growing gap between what children do online, and what their parents think they are doing (Lemish, 2008). With the increased importance of the internet in all of our lives, there is more and more pressure on children to be active on-line, and from a younger age. The dangers permeate almost the entire internet, and change rapidly as the technology evolves. Parents are ill-equipped to protect their children through no fault of their own, but (partly) because while they did grow up in a society where these threats existed, they did not exist in this new form. Today's reductions in barriers to communication have made the problem of protecting children much more complex. Children are often taught not to talk to strangers but with the variety of social interactions available today, teaching a child to block all communications from unknown parties would be challenging to even the most technically minded parent. This is probably undesirable too (Wolak, Finkelhor, Mitchell, & Ybarra. 2008)

This area has understandably received a lot of attention. There is a wide variety of content-control software available to prevent children from accessing illicit material. This mainly works by blocking known URLs, but also by dynamically analysing the content. While by no means trivial, this problem is limited in that it is only the content being sent to the child that needs analysis. These approaches do not apply to two way interactions. Firstly, blocking entire sites/protocols is not necessarily desirable as

some safe use may be allowed (or else the child would be motivated to try work around the blocks). Secondly, the danger a predator poses is not merely displaying unwanted material to the child, but in arranging a meeting outside the parents' control. This can (and may) be done without mentioning anything overtly sexual. Predators are a danger because they can effectively mimic normal child-to-child conversations. If nothing else, one half could be simply copied and pasted from other conversations between actual children. The only difference may be in attempting to meet in person.

This paper discusses the existing approaches to online child protection and the conflicting requirements of the parent and the child in a moderated approach to chat. It presents an idea for a system of anonymous review with various options for added functionality, along with a justification of the system. The penultimate section expands on the two key technical components – pseudonymous messaging and intelligent text analysis.

2. Previous Work

Existing approaches to online child protection typically fall into the following broad categories Block, Review, Filter and Moderate. To enable the reader to better understand the problem domain these approaches and the weaknesses associated with them are examined.

Blocking restricts access to protocols and applications deemed "unsuitable", for example peer-to-peer (P2P) networks or Instant Messaging (IM). Operating in a simple deny/permit fashion can make blocking something of a blunt and unwieldy tool. This lack of flexibility restricts its usefulness only to situations where something must be prohibited.

Reviewing technologies vary in type and application but the core ethos is to allow the parent to monitor the child's activity. Website histories, messaging logs, emails, even full replay of video conference sessions maybe recorded. This may be impractical if the child is an avid internet user or in families with multiple children. Reviewing also suffers from problems of privacy (older children are particularly sensitive about their privacy and may be tempted to circumvent the system) and the generation gap - parents may not be able to penetrate youth lingo and slang.

Filtering may be considered a subset of blocking, usually applied to restrict access to websites considered unwholesome in content. Filters generally consist of blacklisted (or whitelisted) URLs, or dynamic blocking of websites based on content - typically examining sites for a list of proscribed keywords and phrases. Each of these methods suffers drawbacks - blacklisting often involves content labelling, sites labelled as containing certain content are blocked. Labelling is performed by the website operator (who may not be aware of the labelling scheme or may neglect to use it). Some providers of filtering software manually review sites but this is an unscalable approach and the quality of this filtering has been brought into question as has its subjective nature (National Research Council. 2002).

Many internet forums use moderation to enforce rules, edit posts, and ban disruptive users, trolls and spammers. Some child oriented forums, including those of the BBC[30], operate a process of pre-moderation - each message is examined before it is posted. Moderators are trained to screen messages for signs of bullying, harassment, or anything that may result in a child being in exposed to harm. Moderation suffers two key drawbacks - scalability, and the human bias (subjectivity).

The system proposed here addresses the issues highlighted above without sacrificing safety. The child can feel their privacy is being maintained, although messages of concern are being reviewed, the contents of the message will not seen be their parents (thus shielding them from any embarrassment). In this regard the system may be considered similar in fashion to traditional moderation - their messages may be reviewed but not by their parents - but able to overcome its limitations.

3. Conflicting requirements

On-line child protection cannot be solved with technology alone. This paper therefore proposes a system that uses a combination of automation and human judgement to recognise threats. There are many potential pitfalls in trying to solve this problem. One solution might be to give parents comprehensive logs of their child's internet usage. This would be giving them too much information to manage effectively, and would be a tempting target for identity theft. If the parent has the power to

[30] http://www.bbc.co.uk/chatguide/glossary/moderation.shtml.

control exactly what their child does on-line, it is possible they can better protect them, but the controls may be overwhelming. Also, children do not want to have their privacy violated, and will circumvent the system one way or another if it is too invasive. Even if they do not have the level of skill necessary to circumvent the system, they could always spend the majority of their internet time away from the home (at school, library, friends, etc). So some level of privacy for the child is needed. Similarly, all access that can be given and kept "safe" needs to be allowed. It would also be naive to expect children to suddenly migrate onto a new "safe" social network, IM network, etc. Any solution must cater for what they already use.

4. Description

The proposed system uses existing technology as a pre-filtering stage to create a prioritised list of 'suspect' chat conversations. This is subsequently analysed using human judgement via a pseudonymous volunteer who sees an appropriately sanitised version of the data which does not divulge the identity of the child, thus protecting their privacy.

The system works as a software client that can be downloaded and run on any PC. The primary user (presumably a parent of at least one child), installs and sets up the client. There are two stages to the setup. First, the parent must record any sensitive/personal data unique to themselves and their child. This could include names, addresses, email/contact info, credit card number, phone numbers, etc. These will be used to determine when the child may be giving inappropriate information to a stranger, but also when they are inadvertently identifying themselves. The data can be stored as hashed values, or at least encrypted. The second part of the setup is performed in conjunction with the child. The parent reviews or "vets" all contacts the child has (in all applications/platforms/networks). The parent determines which contacts can be considered "safe". This should only include contacts the parent has/can meet in person, or know through some trusted organisation/third-party. At the very least, contacts who neither the child nor the parent has met should be considered unsafe. All contacts are labelled either "safe", "uncertain" or "blocked".

Once setup, the client runs in the background when the child logs onto the computer. The client intercepts all text-based communication protocols before they are presented to the user (the idea is text-based, but see Section 4.8 for voice/video extension). All communications between the child and contacts that have been explicitly labelled safe by the parent continue unimpeded and un-monitored. Any communications to/and from a contact that has been labelled uncertain, or from a new contact get processed. The processing works as follows:

1. All received text from the contact and keystrokes from the child are stored.
2. The text is checked against a list of known problem words/phrases ("sex", "drugs", "would you like to meet", etc.)
3. Other probabilistic analysis is performed (Bayesian network analysis, Gaussian modelling, etc.) looking for indicators of unwanted behaviour.
4. The text is searched for any of the sensitive/personal data entered by the adult. If any is found it is removed and noted.
5. The results from all these tests are combined in a single weighted score.
6. The identifier for the contact (email address, *Skype* name, etc.) is stored as a keyed hash using the child's password (could be their login password).

The processing is done on a section of text of a limited size (a page), and only on communications between the child and one contact. The client stores the processed logs in a list prioritised by the results of the analysis in the above list (top of the list will be the highest match with expected pattern of a predator). As noted in the fourth step, all sensitive/personal data will be removed from the log (can be replaced with a generic placeholder). Rather than just using the data the parent entered in the initial setup, a dictionary of names/local places could also be used to sanitise the logs. Periodically the client will send the logs at the top of the list to another client. Even with the sanitisation of personal data and replacing contact details with hashed values, in order to protect the privacy of the child, the logs will be sent over a pseudonymous network (Kinateder, Terdic & Rothermel, 2005). This allows messages to be passed from client to client, without either being able to discover who the other party is. This works through repeated layers of encryption and routing through various different nodes on the network. The idea is to create a community of effort where parents are reviewing each others children's logs, but in an anonymous manner. The technology used in anonymous communications is described in Section 6.

The receiving client (of the logs) will be administered by another parent. They will be presented with the sanitised logs and asked “Is this something the parent should be concerned with?” The reply can either be a yes/no button, or a scale e.g. from 1 to 10. The second parent can also highlight the text of the logs that is objectionable and a specific reason, e.g. “contact trying to meet child”, “inappropriate sexual content of messages”, “child is revealing personal information/identifying him/herself or home location”. The results are returned to the first parent, again through the pseudonymous network. The first parent will be presented with only those logs (still sanitised) where the reviewer thought there was a problem. The particular application and time/date can be made known to the parent, but the identity of the other contact is still protected by the child’s password. The parent can then decide whether to list the contact as “blocked” preventing any further communication, discuss the matter with their child to determine if the contact can be added to the safe list, or leave the contact “unknown” and continue monitoring.

Figure 1 shows a typical example. In the example, there is a central server called the *Match Maker* that maintains a list of pseudonymous clients that are on-line. An adult can request the details of one or more clients from the *Match Maker*. An alternative method that distributes this information throughout the network is also possible, and discussed in Section 6. The numbered steps in Figure 1 are explained below:

1. Adult1 downloads software client.
2. Adult1 installs client and configures settings (records sensitive data).
3. Adult1 and Child1 agree all safe contacts.
4. Child1 interacts with contacts. Communication with safe contacts continues as normal. Communication with uncertain contacts monitored.
5. Monitoring consists of analysis as described in previous list.
6. Suspect logs are stored in priority list based on results of analysis.
7. Client1 queries *Match Maker* server over pseudonymous network for list of online clients.
8. Client1 selects from list and sends sanitised logs for review.
9. Adult2 checks sanitised logs for undesirable communication.
10. Logs returned to Client1 with rating results.
11. Adult1 takes any necessary action based on feedback.

The most important part of the system is step 9. In order to encourage good behaviour in the clients, each adult will only receive results from the pseudonymous network about their children when they have finished a certain amount of reviews for other clients. This promotes good behaviour. This can be taken a step further by enabling a reputation system at the *Match Maker* server. This would record feedback from adults who found the results they received helpful. The reputation would be tied to the pseudonym and the person behind it would remain unknown, but they could be rewarded by being given a higher priority when requesting (clients for) reviews of their own logs, both in terms of speed and the quality of the rated reviewer.

The following section presents several variations that may improve the overall system, but which should be considered optional as they may have downsides, depending on the implementation details and user requirements.

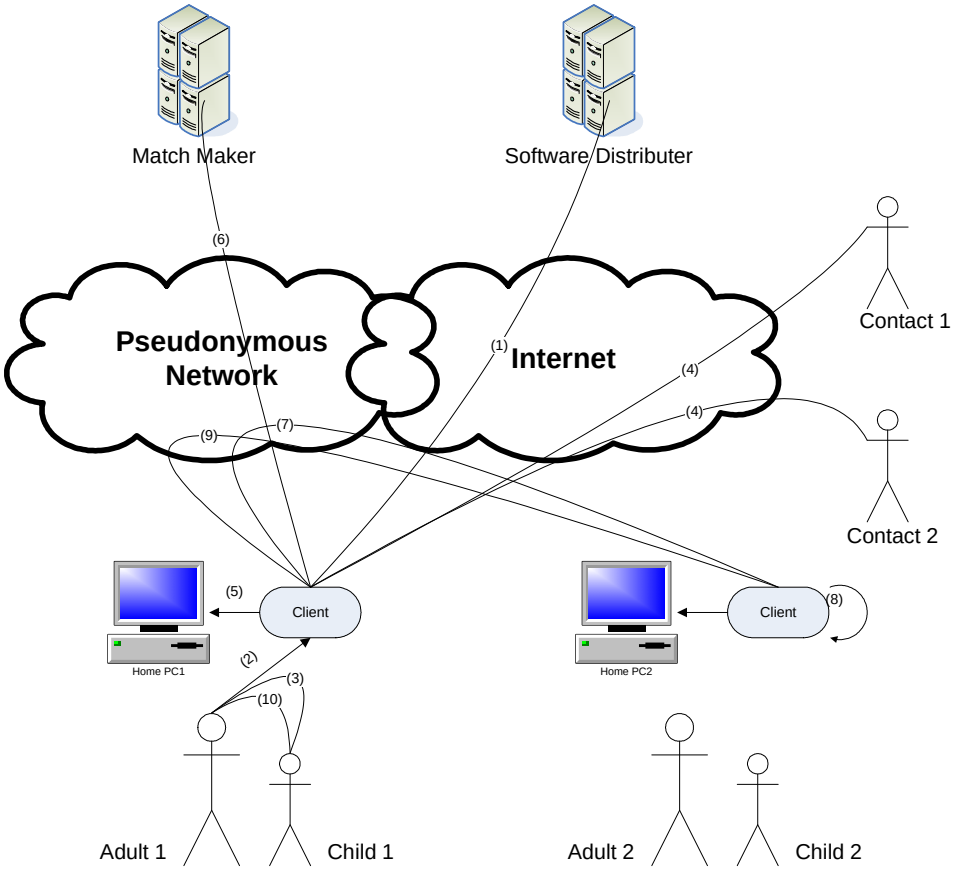


Fig. 1 Communication flow for Pseudonymous Peer Review

4.1. Monitoring all communications

The client could be configured to monitor all communications, including those between the child and safe contacts, but with a higher threshold needed for logs to be stored and sent for review. This negates more of the privacy the child has, but is still a great deal better than traditional monitoring: the contact can still be protected with the child's password, and only logs that have been considered problematic get passed on to the parent. It may be appropriate for younger children.

4.2. Multiple reviewers

The client can send the logs to several other clients for review. This will likely reduce the time it takes for a response, and the redundancy will give the parent more confidence in the results. The downside is the increased burden on reviewers. But given that there may only be a few logs sent, and they can probably be reviewed very quickly, it is likely a good balance can be achieved. For example, the client sends all of the top 3 logs to 3 different clients (sending out 9 in total), but can only review results once they have reviewed 9 other logs. This approach has the benefit of normalising the reviewing process – the effect of a wildly liberal or conservative response would be brought closer to the prevailing societal attitude.

4.3. Instant reaction

Because of the human component of the review process, there will be a delay in the response time. One option that could be considered is that if a given log has a particularly high score, then the client can immediately block communication with that contact and notify the adult, bypassing the anonymous reviewer process. Careful configuration would be needed to avoid too many false alarms. Likewise, the sanitisation could be applied as a filter to all out going communications, not just the stored logs.

This may not be necessary – research indicates (Walsh & Wolak. 2005) that where a victim and sex offender met, the online relationship typically formed over a month or more, from multiple conversations. These would likely have already being flagged, reviewed and reported to a parent.

4.4. Learning behaviour

If the feedback from the reviewers is sufficiently detailed (selection of problem

words/phrases, reasons for choice), then it would be possible for the clients to improve their scoring/prioritisation algorithms. This could either be done locally, or centrally at the *Match Maker* server. Learning at the central server could be very efficient, but would introduce problems with keeping submissions anonymous and being able to trust submissions. Learning locally at the client would not have any privacy problems, but would have a much smaller selection of results to draw from and hence a slower rate of learning.

4.5. Use of dictionary

The client could use a dictionary of common names to sanitise the child's logs. Certain patterns could be identified, such as phone numbers, credit card numbers, postcodes, etc. When the adult enters the home address, a central service might be able to provide a list of all nearby street names, local monuments, and locations that could be used by the child to give directions. This could also be incorporated into the scoring mechanism as well as the sanitisation.

4.6. Re-vetting

A good precaution would be for the adult to periodically re-vet the child's safe contacts. Strangers are not the only sources of potential threats to young children, and asking the child about their relationships with their peers could bring to light early warning signs.

4.7. Impersonation prevention

The times and duration the child has been on-line can be safely stored without fear of privacy invasion. When the adult logs in, a simple calendar with the child's usage can be graphically presented. This prevents the most obvious circumvention method. If there are no usage hours recorded for large amounts of time the parent knows the child was using the computer, then the parent knows the child has circumvented the system (most likely by using the parent's or another un-monitored account).

4.8. Voice/Video chats

As previously described the system is text-based. This can be extended to voice/video, such as commonly found in IM products such as *Skype/MSN*, etc. This requires the use of Speech Recognition Techniques. This technology is currently of limited maturity, but could potentially be employed for keyword spotting. This will slot easily into the proposed system converting the video/audio stream into a text document. This also adds further protection of privacy than would be had by direct monitoring of the video/audio.

4.9 Age of Children

The definition of "a child" covers a large range of ages and maturities. Ultimately the decision to allow a child to use the internet for chat resides with the parent, however a total prohibition would likely lead to using the technology anyway and without parental oversight. The best result is achieved when child and parent have a strong relationship and agree on what is, and what is not permitted. Internet initiated grooming of pre-pubescent children is extremely rare (Wolak, et al. 2008) (Lanning, K. 2001), partly because they are more closely supervised and also because it is difficult to engage them in sexual/romantic conversation because of their immaturity.

This system is best suited to children approaching puberty, they can be informed that (like adults) they are being rewarded with privacy in return for abiding by the rules. As they enter puberty, become sexually aware, and start to desire privacy they will already be familiar with the system. The child would continue to use the system until they reached a suitable level of maturity so that it was no longer needed or they reached adulthood.

Conversations of concern that are sent for moderation would include an indication of the child's age. This would have a bearing on what would be deemed appropriate.

5. Justification

In order for this system to work, active participation of all parties is required. The system has deliberately been designed to encourage good behaviour. The child is motivated to convince the parent that his/her contacts are trustworthy in order to have confidential communications. The parent is encouraged to provide meaningful reviews, in order to get a better reputation, which will result in speedier replies. The process of reviewing can also be of direct benefit to the reviewing adults themselves. It is educational, in that they are made aware of the kind of dangers that exist on-line and

what their child may be exposed to. This better places them to discuss the problems with their child and agree what is safe/acceptable internet usage.

The shortcomings of existing approaches to this problem have been highlighted, and the approach described here overcomes these issues. Conversations, emails and message board postings will be monitored by the emerging generation of intelligent text analysis tools to spot conversations of concern. Only these conversations will be submitted for manual/human review. This process overcomes the scalability issues associated with traditional moderation, vastly reducing the reviewer's workload. The community nature of the approach also negates the cost associated with the traditional moderation model.

Any single moderator reviewing a conversation of concern will be subject to the human condition - bringing their own bias and prejudice into the review process. Using multiple moderators for each conversation of concern will reduce the impact of any one error in judgement. For example say a conversation of concern is reviewed by a conservative individual and someone of a more liberal mindset, an average of their result/feedback will tend towards the centre ground. Here "centre ground" is meant as the general view of society at large.

The moderator will be faced with the hurdle of slang employed by children and may need assistance interpreting the contents of some messages (conversations of concern). It is hoped that such information will be sought from the moderator's own children. The benefits of this will be twofold; the parent will gain a better understanding of trends, challenges and experiences of young people in the modern world helping them to better understand their own children. From reviewing the contents of the messages (conversations of concern) both parent (moderator) and child can learn first hand of the dangers that exist. This will be informative for the parent and can serve as a warning for the child.

6. Components

This section gives a very brief description of how the technical aspects of the system might be implemented. This system relies heavily on the use of pseudonymous communications, along with the dependent technology of Distributed Hash Tables. The first subsection describes the work by Kinateder, et al (2005), the second deals with the work of Stoica, Morris, Karger, Kaashoek & Balakrishnan, (2006) and the third describes how text classifications are best applied to this system.

6.1. Pseudonymous Communications

In order to protect the anonymity of the child, the parent and the reviewer, a communication system with the following properties is required:

- the sender has some ephemeral/indirect knowledge of the receiver, but not their actual address
- the receiver cannot know the address of the sender
- the receiver can reply to the sender
- no observation of the network gives any information about who is communicating with whom

These are all achieved with the use of public key encryption and intermediaries. It is assumed that there is a network of nodes, called *mixes*, capable of processing messages (encryption/decryption) and passing them on. These nodes are not trusted, and while they can disrupt the communications by not participating as expected, it is imperative that they learn nothing by analysing messages as they pass through. It is also assumed that public keys for all parties can be readily obtained (as well as the addresses of the mixes).

However they are stored, the public keys are used to ensure that only the recipient can decrypt the message, but does not help in anonymously delivering the message. Along with the public key, one or more *pseudonym* is stored. These are created by those who want to receive anonymous communications (the recipient) and will be used by a sender. The pseudonym is a sequence of addresses of mixes, but nested in layers of encryption that ensure that only the next mix can decrypt the current layer and they only get the address of the next mix. The further layers of encryption mean that a mix cannot know any further destinations than the next mix, and they do not know any of the previous mixes in the chain since these addresses have been stripped off. Since knowledge of the path of the route is effectively split amongst all these independent entities, it would be impossible to determine the ultimate sender and receiver for a given message without compromising the majority of the network.

There are other ways to increase anonymity, sending messages in batches to confound traffic analysis, sender adding more mixes to the sequence to protect against dishonest recipients, etc. For more details see (Kinateder, et al. 2005). It is also preferable to store the public keys/pseudonyms in a Distributed Hash Table, rather than a single server. This is described in (Stoica, et al. 2006).

6.2. Text Classification

Traditional keyword and regular expression filters are inadequate for analysing IM and chat forum conversations. The new generation of text mining and text analysis tools offer far superior classification abilities. Word frequency, word distance, word pairs (bi-grams), Latent Semantic Analysis, term strength, term frequency-inverse document frequency (tf-idf), and term by document matrices have all proven successful at extracting features from textual sources. These features are subsequently applied to statistical modelling techniques including Bayesian analysis, k-Nearest Neighbour and Support Vector Machines (Tretyakov, 2004) (Conrad & Hunter, 1994). These techniques have proven results in spam detection, document categorisation, authorship attribution and information mining (Aas & Eikvil, 1999).

Classification techniques such as these typically compare new information against known values and categorise it accordingly. For example, a new email will be analysed and the result compared against the result for known spam and ham (legitimate email). This process poses a problem in the child protection domain.

In order to classify conversations as safe or uncertain they need to be compared against known paedophile, and normal chat models. Constructing normal chat models is a trivial task, constructing paedophile (chat) models is not. To construct an accurate model access to paedophile chat logs is required. Law enforcement does not typically share this information freely. Honey trap organisations such as Perverted Justice[31] publish chat logs from sting operations (against paedophiles) on their website. Research (Wolak, J., Finkelhor, D., Mitchell, K., & Ybarra, M. 2008) indicates the in the majority of online sexual predation the offender quickly reveals his true age and intentions – this tallies with the chat logs of Perverted Justice. Pendar's (2007) work shows that Bayesian trigram analysis is effective at detecting grooming (based on Perverted Justice chat logs). The operating practices of Perverted Justice have been called in to question and allegations of entrapment and poor evidentiary quality have been made (Stokley, 2008) (Salkin, 2006). As a consequence of this, the reliability of these chat logs to represent "real world" logs must be considered uncertain. The following proposes another method that may also be suitable - a system of thresholding.

Rather than classifying based on the results of messages matching against two categories/models (normal chat and suspicious), a match against a single model with a threshold value could be used instead. Conversations would be compared against the normal chat model; those matching closely (with a probability above a certain threshold) would be considered normal and not be affected. Conversations poorly matching the model (below the threshold) would be marked as "uncertain" and passed on to the system for evaluation by a moderator. In addition, those conversations falling well below the threshold might be marked for priority moderation/inspection.

In order for the process to work a normal chat model would first need to be trained. The child's online chat activities over a given time would be used to train the initial model. This model will serve as baseline of normal behaviour. Once the training process is complete, online conversations are compared against the model. During the initial training of the model it will not be able to flag problems. As the text classification is just one of several tools used to identify potentially suspicious behaviour, other mechanisms are still available. Key word matches with known general problem phrases and scrubbing of specific personal details is still performed. And to further protect the child (and ensure a "clean" model) the parent can require that no communication with new contacts is allowed for the period of the training. After the initial training period there are likely to be a number of false positive, especially if the training period is brief. If, for example, the child only converses with their peer group during the training of the model, a conversation with a parent or grandparent would be flagged as anomalous. This should be considered a period of normalisation. As the review process records the conversation as a false positive, the text of the conversation can be used to tune the initial model.

[31] http://perverted-justice.com.

7. Conclusions

This paper presents a concept whereby parents can have “conversations of concern” reviewed anonymously by other parents in return for their own actions as a reviewer. The limits of existing technical measures to protect children have been highlighted and it is proposed that the system described here could help bridge the gap through a community approach.

The merits of this idea do not come solely from the technology, but rather from several deliberate reward mechanisms for the users. Children are encouraged to play-by-the-rules and are rewarded with privacy. Parents who are conscientious reviewers will get a better view of the dangers their children are exposed to. The technology provides the means for what would be sensitive information to be shared in a safe way. Great lengths have been made to avoid anything that could be considered censoring by the end-users. All of this is combined to strike the best balance between the child’s safety and their freedom on-line.

References

1. Aas, K., & Eikvil, L. (1999). Text categorisation: A survey. Norwegian Computing Center, Oslo, Norway, http://citeseer.ist.psu.edu/aas99text.html

2. Conrad, J. G., & Hunter, M. U. (1994). A system for discovering relationship by feature extraction from text databases. Annual ACM Conference on Research and Development in Information Retrieval, 1994

3. Kinateder, M., Terdic, R., & Rothermel, K. (2005). Strong Pseudonymous Communication for Peer-to-Peer Reputation Systems. ACM Symposium on Applied Computing, 2005. http://portal.acm.org/citation.cfm?id=1067033

4. Lanning, K. (2001). *Child molesters: A behavioral analysis* No. Fourth Edition)National Centre for Missing & Exploited Children.

5. Lemish, D. (2008). Generation Gap? ’Online Gap’ Widens Divide Between Parents and Children. Science Daily. http://www.sciencedaily.com/releases/2008/02/080204143203.htm

6. Salkin, A. (2008). As perverted-justice.com battles web pedophiles, some raise concerns over its tactics. International Herald Tribune. http://www.iht.com/articles/2006/12/13/news/justice.php?page=1

7. Stoica, I., Morris, R., Karger, D., Kaashoek, M.F., & Balakrishnan, H. (2001). Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications. Proceedings of the 2001 ACM SIGCOMM Conference. http://pdos.csail.mit.edu/papers/chord:sigcomm01/chord sigcomm.pdf

8. Stokley, S. (2008). ’To catch a predator’ sex stings net mixed results. The Press Enterprise. http://www.pe.com/localnews/inland/stories/PE News Local R dateline28.6b3814.html

9. Tretyakov, K. (2004) Machine Learning Techniques in Spam Filtering. Institute of Computer Science, University of Tartu.

10. Walsh, W., & Wolak, J. (2005). Nonforcible internet-related sex crimes with adolescent victims: Prosecution issues and outcomes. *Child Maltreatment, 10*(3), 260-271.

11. Wolak, J., Finkelhor, D., Mitchell, K., & Ybarra, M. (2008). Online "predators" and their victims myth, realities and implications for prevention and treatment. *American Psychologist, 63*(2), 111-128.

12. Wolak, J., Mitchell, K., & Finkelhor, D. (2006). Online Victimization of Youth: Five Years Later. http://www.missingkids.com/en US/publications/NC167.pdf

4. STEGANOGRAPHED CUSTOM EMOTICONS

Digital Forensics is a branch of forensic science that deals with investigations based on evidence stored on digital media, for example recovering deleted files that are indicative that a computer was used in a crime. This chapter describes how a unique image can be used as a 'digital fingerprint', left on the computers of individuals a user has been chatting with.

Instant Messaging (IM) is a form of computer mediated communication, typically text based, though many IM platforms include the option for audio and video. Traditionally bound to the computer, these platforms are now commonly available on phones and portable devices.

In 2004 IM was the fourth most common activity among children on the Internet [1], in 2008 it was the second most popular activity[32] [2], 61% of children regularly use IM [9].

A common feature of text based communication is the emoticon (or smiley). These allow the user to increase the information content [3] of the conversation by introducing body language, in the form of a crude facial expression. Most IM applications, some chat rooms and message boards include a set of default emoticons covering the basic emotions (happy, sad, angry, embarrassed, etc). *Windows Live Messenger*, also called *MSN Messenger* (sometimes abbreviated to simply *MSN*) allows users to create their own custom emoticons.

Using steganography techniques to create pseudo-unique images and taking advantage of the design properties of the underlying protocol it is possible to leave a 'digital fingerprint' (a Steganographed Custom Emoticon (SCE)) on a receiver's computer without their knowledge and without employing any 'hacking' techniques. This emoticon can later be recovered and used in non-repudiation, for example a groomer denies conversing with a child, if an SCE is found on his computer this can be used as circumstantial evidence in the case against the groomer.

[32] The age range of 2004 is 10-17, for 2008 the range is 10-15.

This chapter shall examine the background of emoticons, the popularity and risks of IM and how Internet criminals are apprehended. It will then describe the process of designing an SCE, the method of operation and a justification of the concept.

4.1 Background

Face-to-face communication typically consists of spoken words, the tone of delivery and non-verbal communication (body language). Text based communication loses a significant portion of the information content associated with face-to-face communication [3]. *Scott Fahlman* is credited for suggesting :-) [colon, hyphen, bracket] to express a joke or happy emotion on a *Carnegie Mellon* bulletin board in 1982 [4], the idea was not new - alphanumeric combinations were commonly used to express emoticons.

The increase in online text based communications saw ASCII art (creating an image from the 95 printable ASCII characters) find mainstream acceptance. Many applications now intercept the character strings and display an image. The increase in computing power, Internet access speeds and rich media has led to graphical emoticons in many applications.

4.2 Catching Internet Criminals

Internet criminals, including child molesters are typically caught by one of three methods – through the identification of their IP address [16] [14], disclosure of information that identifies them in the real world [12] [13], or they come to the attention of law enforcement for other reasons and their Internet crimes are uncovered [15].

All Internet traffic has a source and destination Internet Protocol (IP) address. The *Internet Assign Numbers Authority*[33] (*IANA*) oversees global IP address allocation. *IANA* provides blocks of IP addresses to Regional Internet Registries - *Réseaux IP Européens Network Coordination Centre* (RIPE NCC, RIPE)[34] in the case of the UK. *RIPE* assigns IP address blocks to local Internet registries or Internet Service Providers (ISPs).

[33] http://www.iana.org/ (accessed 1 July 2010).
[34] http://www.ripe.net/ (accessed 1 July 2010).

IP addresses are frequently recorded in server traffic logs and can be captured from live traffic with protocol analysers (packet sniffers). Internet criminals are traced through their IP address via their ISP's subscriber database.

Section 11 of the Anti-terrorism, Crime and Security Act 2001 [5] defines a code of conduct for ISPs to retain subscriber details for 12 months.

Avoiding detection by IP address is a trivial affair – using a public terminal (café, library, 'wardriving'), or hiding their IP address behind a proxy or other 'anonymising' (privacy) service.

Criminals may divulge personally identifying information, this could be revealed to their victim or another individual (and later relayed to law enforcement), or they may disclose it to law enforcement in a sting operation.

In the course of an investigation law enforcement may uncover details of an individual's involvement in an otherwise unknown Internet crime.

4.3 MSN

MSN is often used as a vague term to describe a number of applications and services. Here the term is used to describe the system as a whole.

The *.NET Messenger Service* (formerly *MSN Messenger Service (MSNMS)*) is an instant messaging and presence system, powered by the *Microsoft Notification Protocol (MSNP). Microsoft* released the protocol to the *Internet Engineering Task Force (IETF)* in 1999 and allows third party clients to use the service [6] [7].

Windows Live Messenger (formerly *MSN Messenger*) is a client application from *Microsoft* that uses the *.NET Messenger Service* to communicate. Numerous third party clients exist that port access across operating systems and hardware platforms.

In 2006 Europe had 82 million IM users, 60% of which use MSN [8].

MSN supports a standard set of built-in emoticons. When sending these default emoticons the application transmits only the character sequence (string) corresponding to the emoticon. The character sequence is referenced to an emoticon. Both the transmitting and receiving client show their own copy of the emoticon (the emoticon image bundled with the application). For example *user A* enters :) [colon bracket], the transmitting client (application) recognises the string as an emoticon 'reference' and displays the emoticon to *user A*, typically a smiley face. The client will also send the string to *user B.* The receiving client (*user B)* will recognise the string reference and also display an emoticon. When standard emoticons are used only the reference character sequence is sent, not the emoticon image itself. If different clients are used the users may be shown different images (though these will depict the same emotion and may look very similar). Figure 8 illustrates the difference between default and custom transfer.

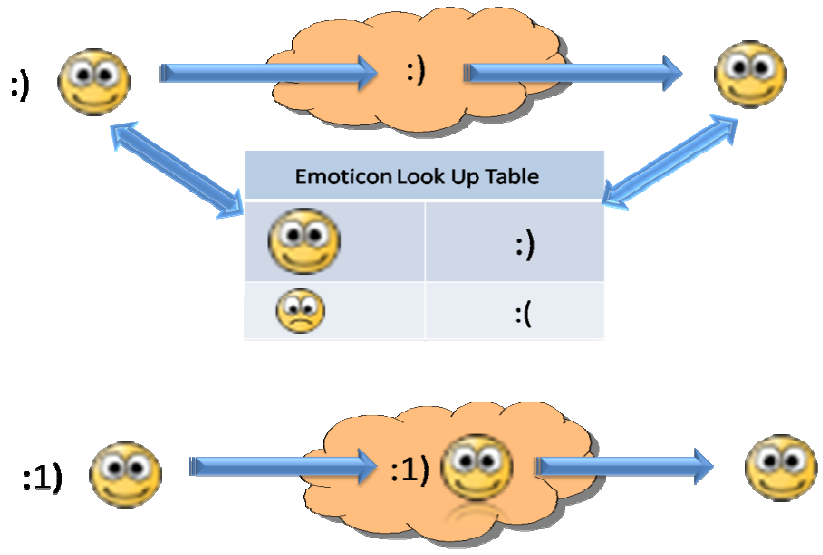


Figure 8 - Default vs custom emoticon transfer

The process for custom emoticons is different. When creating a custom emoticon the user loads an image into the application and assigns it a character sequence (this cannot be a sequence assigned to a standard emoticon). When the user enters the (custom emoticon) character sequence, both the character sequence and the emoticon image are sent to the receiver (the receiver cannot look up an image it does not have). Both users see the same image.

The receiving client assigns the image a random filename and stores it deep in the file system.

The receiver can save the SCE from inside the application or may find it (applications exist to recover emoticons). If the receiver tries to the copy the emoticon, the image file contents change (not visually) as a side effect of the application compressing the image to use as an emoticon. This allows the system to be used in non-repudiation.

When sending a file or image (as a file), the transfer has to be accepted by the user. Emoticons and user icons are transmitted automatically in the background without user interaction (drive-by download).

4.4 'Steganographing' and Detecting Custom Emoticons

Any unique or pseudo-unique image could be used in the system. Using an innocuous image will increase the efficacy of the system, by not arousing the interest of the recipient, a minor alteration to a default or plain emoticon would attract the least interest.

For the system to operate each enrolled user requires a unique emoticon. This is achieved using techniques from steganography – the science of hiding information is plain sight, so that only an informed individual can detect the message. This process alters the emoticon in a way that makes it unique[35] but detectable and verifiable.

The emoticon is 'steganographed', made unique by altering the 24 bit value of a single pixel in the image. The colour of each pixel is determined by the value of the three colour channels - RGB (red, green, blue). Each channel has 256 shades (0-255), the product of the channels (256*256*256) provides the possibly of 16.7 million colours. This can be extended using the alpha (a) channel; this determines the opacity of the pixel and extends the entropy to 32 bits. A minor change in any of these channels is undetectable to the human eye.

[35] There are hundreds of possible emoticons (see http://www.sherv.net/Free-MSN-Emoticons.html (accessed 02 Sept 2010) for some examples). Each emoticon can by 50x50 = 2500 pixels, any of the pixels can be doctored in one of more channels by at least +/-5. Though not mathematically unique the number of possible source emoticons, number of pixels per emoticon and colour range of pixels mean the chance of two individuals independently creating the same image is improbable enough to be reasonably termed unique.

The emoticon is identified by hashing. A hash is a one-way mathematical function that generates a fixed length integer irrespective of file size. The nature of hashing means that a minor change in the data results in a dramatic change in the hash. Therefore a minor change to the colour and/or opacity channels of a pixel, though not discernable by eye is readily revealed by hashing. Figure 9 illustrates the source emoticon and a steganographed derivative and the differences in the hashes.

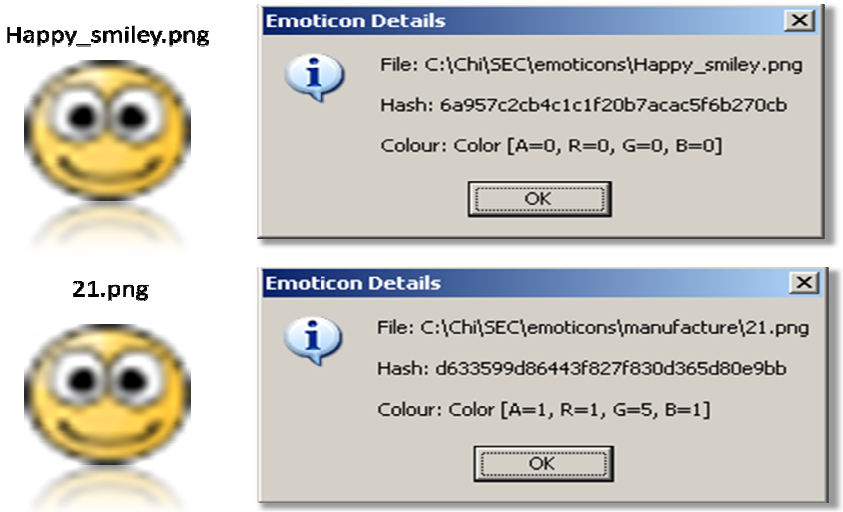


Figure 9 - Source and steganographed emoticons and hashes

4.5 Method, Experimentation and Conditions

A demonstrator was created as a proof of concept exercise. An application was written to amend the aRGB channels of a pixel. The source emoticon image (a .PNG file) was chosen from the default set of emoticons that ships with the open source messenger client *aMSN*[36]. The application makes a change of +/-1 (in the range up to +5 if the original value is 0-127 or up to -5 if the original value is 128-255) to one of the aRGB channels and saves the new image. It then makes a change in the subsequent channel and saves the new images, etc[37].

Altering each channel +/-1 to a limit of +/-5 gives a total of 625 different images (5*5*5*5 = 625), per pixel changed. Emoticons can be up to 50x50 pixels.

[36] http://www.amsn-project.net/ (accessed on 5 July 2010).
[37] Full source code can be found in the appendix (see Image Fiddler).

Each time the image is added to the application and turned into an emoticon, compression changes the file structure (and the hash). In order to determine the hash of the received emoticon it (the SCE) needs to be transmitted to and hashed at a control receiver. This process is shown in figure 10. In a real world production context, the SCE would be shipped to the user in a custom MSN client downloaded by a user (parent) when enrolling in the system. The user's registration details would be stored in a managed service (database) along with the SCE hash.

During experiments the MD5 hash was used, in a production environment the cryptographically harder SHA-2 would probably be more suitable.

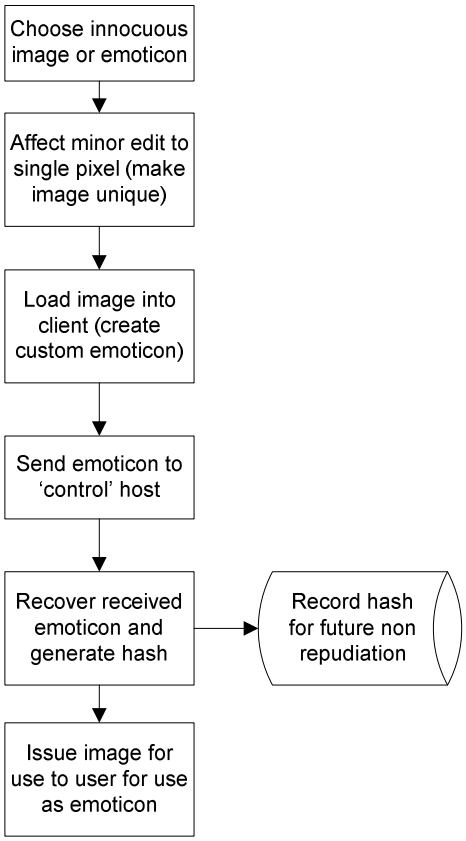


Figure 10 - Emoticon manufacture and issue

If a computer is seized and examined in the course of law enforcement investigations, custom emoticons found on the suspect's computer can be checked against a registry of SCE, i.e. those issued to users. If a match is found in the database i.e. the emoticon has been issued to a child, this is circumstantial evidence that the suspect has been communicating with the user. If the suspect denies such communication has taken

place the uniqueness of the emoticon (the hash) can repudiate this claim. The system has been described in terms of groomers and children, but could function in any setting where communication via MSN has taken place and this communication is been repudiated.

Figure 11 describes the process surrounding the recovery of the emoticon.

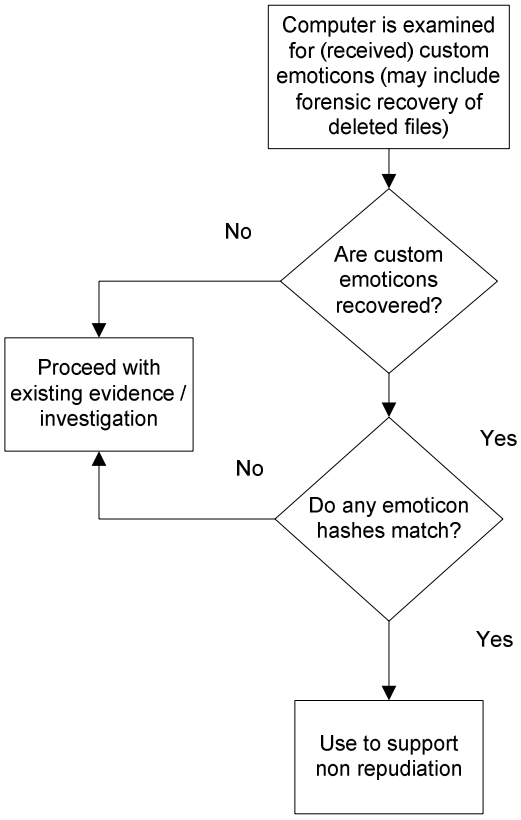


Figure 11 - Emoticon recovery process

Numerous third party clients are available, not all of these implement the complete feature set, including custom emoticons. In experiments both *Windows Live Messenger*[38] and *aMSN* have been observed to receive and store custom emoticons.

The penultimate stage of phase one (emoticon manufacture and issue – see figure 10) is to recover the emoticon hash (from a 'control' host) and record it for future use. A second complimentary application[39] was written to support the concept demonstrator. When the emoticon is received (at the control host), the operating system creates a

[38] http://explore.live.com/windows-live-messenger?os=other (accessed on 06 July 2010).
[39] Full source code can be found in the appendix (see Hash Checker).

file. The application monitors the directory where the emoticon is created. The file creation triggers the application to hash the emoticon (file) and enters the hash into a database form. Additional fields exist for user details. This process can also be performed manually from a GUI. A hash (from a recovered SCE) can be entered into a database form to verify if it has been issued to a user, as in figure 11. Data verification and error checking is performed at each stage to ensure an emoticon is not issued to multiple users.

Experiments were conducted across two laptops, one designated as belonging to a child (the user / sender of the emoticon) – Laptop1, the other as belonging to an unknown recipient – Laptop2. A sample of emoticons were hashed and their aRGB values recorded – this step is unnecessary and was performed for completeness. The emoticons were loaded into the application on Laptop1 and used in conversation (sent to) Laptop2. The emoticons were copied from Laptop2, their hashes and aRGB values recorded. The emoticons were then deleted from Laptop1. To simulate the recipient reusing an emoticon, the emoticons (received on Laptop2) were loaded into the application on Laptop2. These were sent to Laptop1, recovered and the hashes checked to ensure they had changed, as would be necessary in a non repudiation system. A table of the results can be found in *Appendix D*. Hashes and channel values where examined with the applications described, data streams and packets were examined using *WireShark.*

4.6 Justification and Discussion

Outside of the lab in a production environment, the author envisages this would be provided as a subscription service. A user, typically a parent, would register and download a custom application for the MSN service. The SCE would be embedded in the custom front end. The SCE would be sent each time a conversation was started with a new user, prefixed to the opening chat. This would overcome the need for the child to manually send the emoticon.

As discussed in chapter 2 (IdM), the Internet was not designed to support identity claims, neither does it have a mechanism for non-repudiation. Digital signatures, part of the Public Key Infrastructure (PKI), are used online to verify the authenticity and source of information and in non-repudiation. *Files and messages are digitally signed*

with a private key (only available to the source) and verified with a public key (available to everyone).

Steganography is the science of hiding information in plain sight. In this system, a minor change in the aRGB channels of a pixel is used to create a unique custom emoticon.

A function of the *.NET Messenger Service* is the drive-by download of custom emoticons to recipients' computers, this places custom emoticons on the receiver's computer without user action. Using steganographic techniques the custom emoticon is altered to make it uniquely identifiable. This allows it to serve as a 'digital fingerprint' that provides circumstantial evidence that a conversation took place.

There is scope to argue the SCE could be planted (deliberately copied from the receiving computer to another), so cannot conclusive proven guilt. This is true, like deoxyribonucleic acid (DNA) and other types of trace evidence, the SCE could be planted. Discovery of the SCE, like the discovery of trace evidence does not prove anything – it is circumstantial evidence, it *is* strongly indicative that communication has taken place. *But this does not mean the conversation was untoward.* If a forensic examination fails to recover an SCE, this is not proof that the communication did *not* take place – the recipient may have used a different computer, or a client that doesn't support custom emoticons.

It can be argued that recipients could delete[40] all custom emoticons from their computer, even using software conforming to DOD 5220.22-M[41] to securely purge the emoticons. Criminals are frequently caught from fingerprint evidence, despite the knowledge that gloves would prevent this. Whilst custom emoticons can be removed, a significant number of individuals will be unaware of their existence, lack the technical proficiency to remove them, or do not believe they will be caught. The threats from viruses and malware on the Internet are well known, yet inadequately patched and protected computers continue to be compromised in large numbers [15]

[40] Providing the disk space has not been over written, deleted data can be recovered using carving tools.
[41] http://www.dtic.mil/whs/directives/corres/html/522022m.htm.

[16], due in part because of the same reasons (lack of knowledge, inadequate technical proficiency, belief they will not be affected).

Parents could ban their children from using instant messaging or monitor their children's communication *(existing online child protection methods are more fully discussed in later chapters).* This may be effective with younger children but older head strong teens will find alternate communication channels or access points if communications are prohibited. Research [9] suggests that parents are unable to decipher youth Internet lingo, rendering monitoring a largely ineffective activity.

This approach has limitations and requires certain conditions to exist; custom emoticons are not supported by all clients, and emoticons can be disabled (by default they are enabled), *beyond this a user could be on a public terminal (a library or Internet cafe) or using a live Linux distribution. Beyond this IM maybe in decline due to the rise of Facebook messaging.* A system could be designed to enforce real world identity registration, log all conversations to a server and have all communications digitally signed. Such a system would be an expensive and invasive product that offers the user no additional functionality and has serious privacy implications. The system described here has been developed around an existing system and its limitations.

Despite these limitations the author feels the benefits of such a system make it a worthwhile endeavour. With the existing infrastructure in place development costs would be minimal. With no tracking or remote retrieval capabilities the SCE is no threat to privacy. The benefit of this system is threefold, first is the ability to provide law enforcement additional evidence to support cases against groomers and child molesters. Second, it may also provide a deterring element, dissuading groomers from contacting children to begin with. Third and more abstract is parental reassurance. Furedi [9] argues that fuelled by media scare stories, parents have become excessively fearful for their children's safety and have developed an overbearing approach that is retarding child development. Knowing a system such as this exists and may have a deterring effect on abusers, could reassure parents, bringing their perception of risk more in line with actual risk and providing peace of mind. *The concept of risk perception verses actual risk is explored in later chapters.*

4.7 Conclusion

Instant Messaging is a popular Internet communications tool; it allows one-to-one and group conversations to occur where a relationship already exists. Initially text only, IM has become a rich media experience with support for voice and video.

IM is popular among children and teenagers, who often use a lingo impenetrable to their parents [10]. More grooming and Internet initiated sex crimes begin on IM that on other technologies [11].

MSN is a popular messaging platform with tens of millions of users in Europe alone. Although *Microsoft* controls the authentication process and setup and control traffic, the protocol is in the hands of the IETF and third party clients are able to access the system.

This chapter has described a method for placing a uniquely identifiable emoticon on the computer of anyone who chats with a user of the system. The latter recovery of this emoticon can be used to reasonably conclude the recipient of the emoticon has been in communication with a specific user.

The system is not a model solution it has been developed around the limitations of existing technology. However the author feels the system is worthwhile, it can help law enforcement in their pursuit against groomers, it may act as a deterrent against grooming and it could help with parents' peace of mind.

4.8 Business Impact

This work demonstrated the feasibility of exploiting the behaviour of the *.NET Messenger Service* to leave a 'digital fingerprint' in the form of a 'unique' emoticon on the receiving computer. The method use stegangraphic techniques to avoid the arousal of interest or suspicion.

The approach was warmly received but is a high risk to develop into a managed service for consumers – it would require considerable marketing and PR involvement,

does not ‘fit’ well with BT’s core business interests and would be difficult to monetise.

BT showed considerable interest in patenting the idea, but it was felt that publishing the information and allowing third parties to develop the concept would provide value to the BT brand and reputation by showing that BT was active in the problem space and eager to share its findings in the public domain.

4.9 Steganographed Emoticon Report
The following report, has been circulated within BT. It expands on ideas presented in the chapter.

Steganographed Custom Emoticons – 'Digital' Fingerprints in MSN Messenger

Abstract
Instant Messaging (IM) is a form of computer mediated communication in which (usually) two or more individuals converse in a text based format via the Internet. Modern IM is a media rich environment with support for graphics, audio and video content. Emoticons, also called smilies are commonly used in IM and a limited set is often supplied as standard with applications. IM applications may permit users to create their own emoticons (custom emoticons) and send these to other users during conversation. This paper proposes a way to discretely leave a 'digital fingerprint' in the form of a steganographed image onto a host computer during an IM chat session. This image can be recovered later and used in non-repudiation to support law enforcement cases against Internet child groomers. The approaches and methodology discussed in this paper are based around the *Microsoft MSN* protocol.

Keywords
MSN, MSN Messenger, Custom Emoticons, Non Repudiation.

Authors:
Chi Durbin
Mark Pawlewski
David Parish

ISSUE Issue 1.0
DATE 1st October 2009

Document Information

Title	Steganographed Custom Emoticons – 'Digital' Fingerprints in MSN Messenger
Owner	Chi Durbin
Author(s)	Chi Durbin, Mark Pawlewski, David Parish
Location of electronic copy	
Location of paper copy	None
Change Authority	

Change Control

Distribution

Issue	Author	Date	Details of Change

Confidentiality Statement

Copyright

1 INTRODUCTION 76
2 EXISTING CHILD PROTECTION METHODS 76
3 CATCHING INTERNET CHILD SEX OFFENDERS 77
4 MSN 78
5 CUSTOM EMOTICON MANUFACTURE AND DETECTION 79
6 JUSTIFICATION 81
7 CONCLUSION 82
8 REFERENCES 84

INTRODUCTION

Instant Messaging began as a method to engage in text based 'conversations' over a network - *User A* would type a message that would appear on the screen of *User B, B* could reply and so on.

IM is a popular communication media with children, and research indicates children are more likely to be groomed using IM than social networking platforms such as *MySpace* (Wolak *et al.* 2008). Various independent software vendors have tools that claim to make IM environments safer for children – these are discussed later. None of these tools leave evidence on offenders' computers that can be used later to support legal action. This paper describes a technique whereby forensic evidence is left on the computer of anyone chatting with a child.

Much information is lost in text based communication, not least is the loss of emotion portrayed in the face (Hancock *et al.* 2007). Many users use emoticons to help represent their emotional state and add to the informational content of the conversation. Although using alphanumeric combinations to represent words or emotions was not new, *Scott Fahlman* is credited with first suggesting the use of :-) [colon, hyphen, closed bracket] to represent a happy emotion or joke on a Carnegie Mellon bulletin board in 1982 (Kharif. 2001). This has since been extended in many ways.

As IM gained popularity ASCII art received broad mainstream acceptance. Many applications now intercept the character strings and represent these as images. As IM became a mainstream communications channel especially on the Internet, developers added rich media tools to IM applications including support for icons. Icons expressing emotions for example *smiley faces* are de facto referred to as 'emoticons' (this paper assumes the use of this term). As Web 2.0 swept the Internet, developers included the ability for users to create their own custom emoticons.

The method described herein takes advantage of the *.NET Messenger Service* (formerly *MSN Messenger Service* (MSNMS), often referred to simply as MSN) and the *Microsoft Notification Protocol* (MSNP) with regard to the way in which it handles and transmits custom emoticons. The technique allows an MSN user, in this case a child, to place a unique, identifiable, custom emoticon on the machine of every person that they are interacting with via MSN text chat. This is effectively their 'digital fingerprint', and can later, if necessary, be recovered and used in non-repudiation in support of law enforcement activities. The important point about this fingerprint is that it occurs as an intended function of how the MSN protocol operates rather than a security exploit or vulnerability.

EXISTING CHILD PROTECTION METHODS

There are numerous technical measures for protecting children on the Internet, these include products by *Net Nanny*[42], *CyberPatrol*[43], *CyberSentinel*[44] and others. These typically take the form of filters, blocks, logging and monitoring.

Crisp Thinking[45] has developed a tool described as an *Anti-Grooming Engine (AGE)* which builds a behavioural profile of the child's chat. Bayesian inference (Crisp Thinking. 2008) is used to evaluate the conversation for sexual content, punctuation, levels of aggression, typing speed, sentence length and vocabulary, and compares this with known profiles for real-life groomers and children. *Crisp Thinking* claim 98.4% accuracy in independent tests (M2

[42] http://www.netnanny.com/ (accessed 9th September 2008).
[43] http://www.cyberpatrol.com (accessed 9th September 2008).
[44] http://www.cybersentinel.co.uk (accessed 9th September 2008).
[45] http://www.crispthinking.com (accessed 9th September 2008).

Presswire. 2008), but beyond these headline figure results, test data and methodology are not in the public domain[46].

Some child protection software filters IM communication in much the same way as websites are filtered. Children are prevented from revealing personal information (address, phone number etc). A list of proscribed terms common in sexual conversation and harassment are blocked, and lists can often be enhanced or extended based on the level of protection desired. Filtering is not an ideal solution; many sexual terms also have legitimate non-sexual application. Sexual innuendo and slang terms frequently evade filters. Many terms commonly have different meanings depending on the context in which they are delivered.

Many Internet security suites log all IM activities, contacts and conversations to be reviewed parents. For parents with multiple children or whose children are avid IM users this may quickly become an unsustainable activity.

Blocking of IM, either preventing the application executing or accessing the network is another less flexible option. This approach may be appropriate with very young children, but as many messaging applications have web interfaces this is easily bypassed.

The *Internet Watch Foundation*[47] (IWF) maintains a list of child abuse websites; the list is updated twice daily and contains 800 – 1,200 live sites. About 50 URLs are added daily (Ozimek. 2008).

BT CleanFeed is an ISP level filter used by major ISPs. *CleanFeed*[48] is a two stage filtering process – if a request is made for information hosted at 'suspect' IP address it is routed to a HTTP proxy. If the information is not hosted at a suspect IP address it is sent to the destination IP address as normal. Suspect traffic is examined in detail at the proxy and compared to the IWF blacklist (Clayton). Requests for child abuse content are blocked, other requests are permitted. *CleanFeed* only processes UK traffic and can be circumvented using international proxies. Nevertheless, it does provide a level of protection to the general user who is not trying to intentionally bypass it. It also removes the 'accidental access' defence from people found to have accessed child pornography.

Protecting children on the Internet is not simply a matter of software and technology, or unilateral action. Parents must adopt a holistic approach and work to educate their children about potential threats and risky behaviours.

Older children and teenagers value their privacy and object to having their communications blocked or monitored. They may also have sufficient technical proficiency to bypass these measures. While younger children may more readily accept this, they require closer supervision as they are more naïve. Mitchell *et al* (2007) contend that parents should engage with children, establish rules and help them understand the dangers, but that banning Internet chat is probably ineffective with teenagers (Mitchell et al. 2001).

CATCHING INTERNET CHILD SEX OFFENDERS

Currently there exists two distinct methods for law enforcement agencies to trace sex offenders on the Internet. These are via the identification of an IP address or through the revelation of personal information (either in a 'sting operation' or passed by a victim/on behalf of a victim). This personal information might be gleaned across a variety of sources where the offender has used the same user name or email address.

[46] The authors have asked to see information.
[47] http://www.iwf.org.uk/ (accessed 2nd September 2010).
[48] The exact design of CleanFeed is not in the public domain.

All Internet traffic requires a source and destination IP address. *The Internet Assigned Numbers Authority* (IANA)[49], manages the global IP address space. IANA works with Regional Internet Registries - *Réseaux IP Européens Network Coordination Centre* (RIPE NCC)[50] in the case of the UK – to assigned IP address blocks to local Internet registries or Internet Service Providers (ISPs). ISPs assign IP addresses to individual consumers and organisations.

IP addresses may be recorded in traffic logs or captured by protocol analysers often called 'packet sniffers'. Many Internet criminals are traced via their IP address, as ISPs can map the address to subscriber information records. Although criminals can be identified via their IP address, the technique is unreliable as suspects may use a public terminal (library, Internet café, 'Wi-Fi hotspot'), conceal their address behind a proxy or make use of a range of privacy techniques that are readily available.

Suspects may also be traced from personal information they disclose either in conversation or information disclosed in their profile such as name, telephone number or address. Once a 'real life' suspect has been located, law enforcement can begin examining their computer for evidence of grooming (and other criminal activity). It is this stage where the (recovered) custom emoticon becomes relevant.

MSN

The *.NET Messenger Service* colloquially referred to as MSN is a messaging and presence service that was released to the *Internet Engineering Task Force* (IETF) in 1999 (Microsoft. 1999, Movva & Lai. 1999]. The service can by accessed using *Microsoft's* own client *Windows Live Messenger* (formerly *MSN Messenger*) or by third party application. As of 2006 Europe had 82 million IM users, 60% of whom use MSN (comScore. 2006).

MSN supports a standard set of emoticons. These are handled by passing the appropriate character sequence between clients. The character sequence is referenced (by the application) to an emoticon and each client will show its 'own' copy of the emoticon. For example *user A* enters :) [colon bracket], the client recognises the string as a emoticon 'reference' and displays the emoticon to *user A*, typically a smiley face. The client will also send the string to *user B.* The remote client (*user B)* will recognise the string reference and also display an emoticon. When standard emoticons are used, only the reference character sequence is sent not the actual emoticon.

For custom emoticons the situation is slightly different. If a user types the character sequence representing their unique custom emoticon, their client (application) will transmit the key sequence and the emoticon - the receiving computer will not have a copy of that unique emoticon. The custom emoticon is then embedded on the receiver's machine and behaves like other standard emoticons for subsequent interactions. If the custom emoticon is unique to the sender this provides strong evidence that there was an interaction between the two parties.

An obvious potential flaw in this system is that the receiver of the custom emoticon could potentially re-send the emoticon onto a third party which would erroneously imply that the child had also been interacting with this third party. However, this is actually not an issue because of a side effect of compression which changes the hash of the (resent) emoticon even though it may look the same.

[49] http:// www.iana.org/ (accessed 20th August 2010).
[50] http://www.ripe.net/ (accessed 20th August 2010).

CUSTOM EMOTICON MANUFACTURE AND DETECTION

From the technical standpoint the system will function with any emoticon, nevertheless it is preferable to use an innocuous image in order to avoid attracting attention and to minimise the chances of the recipient attempting to identify and delete it from their machine.

The aim of generating the custom emoticon is to make it unique to the user, yet relatively innocuous in its appearance. The approach borrows the idea from steganography - the science of hiding information in plain sight. An unsuspecting recipient may examine the emoticon and be presented with seemingly normal image. An informed user will however be able to extract additional 'secret' hidden information and identify the emoticon as being unique to the sender.

This differs from cryptography in that cryptography will be apparent to any user – the user may not be able to decipher the encrypted message (only those possessing a key are able to access the information), steganographed information is available for all to see but one must know where to look. As a real world example, cryptography may be compared to a safe – everyone can see the safe but only those with a key or combination have access. Steganography is comparable to a hidden compartment or false wall – unless one knows where to look, one will find nothing.

The unique custom emoticon is created by altering the colour value of a single pixel in a 24bit image. This subtle change is enough to make it unique yet without changing its overall appearance. Pixels have three colour channels (red, green and blue) each with 256 different shades, these combine (256*256*256), to provide 16.7m possible colours for each pixel[51]. Therefore a minor change in any channel will change the pixel (and thus the image) by an amount that can be detected by a computer but not the human eye. The product of this change is a unique[52] image.

When an emoticon becomes embedded on a recipient's computer, it is not a straightforward task to find it. It is buried deep in the file system and does not have a readily identifiable filename. If a situation arises where the presence of the emoticon on a recipient's machine is required for evidence, then it is of course necessary to be able to easily identify the emoticon file.

In order to facilitate identification, the system generates a hash of the emoticon file prior to sending it. This hash can then be used to uniquely identify the emoticon. A hash is a one-way mathematical formula used to generate a unique fixed-length integer value regardless of the size of the source file or message. A minor change to a file or message will completely alter the hash value. The hash value is therefore unique for that particular emoticon and the emoticon can be identified by generating hashes for all emoticon files on the recipient's computer and finding the file that yields the expected hash value[53]

There is an additional complication in generating the hash value in that the client application (*Windows Live Messenger*) compresses the image used when creating the custom emoticon, thereby changing the hash value from that of the original image. It is therefore important to generate the hash from the file that the recipient will receive rather than the hash of the original emoticon image. In practice the best way to do this is to send the emoticon to a 'control' machine and then generate the hash of the file received on that machine. If the

[51] In tests the authors used images with opacity support increasing the entropy to 32 bits.
[52] The number of possible source emoticons, number of pixels per emoticon and colour range of pixel means the chance of two individuals independently created the same image is improbable enough to be termed unique.
[53] Popular hash functions include MD5 and the National Security Agency (NSA) developed SHA-1.

receiving user attempts to recreate and resend the emoticon, as noted above, the image will be recompressed and the hash will change.

The process for generating the unique custom emoticon is given in Figure 1.

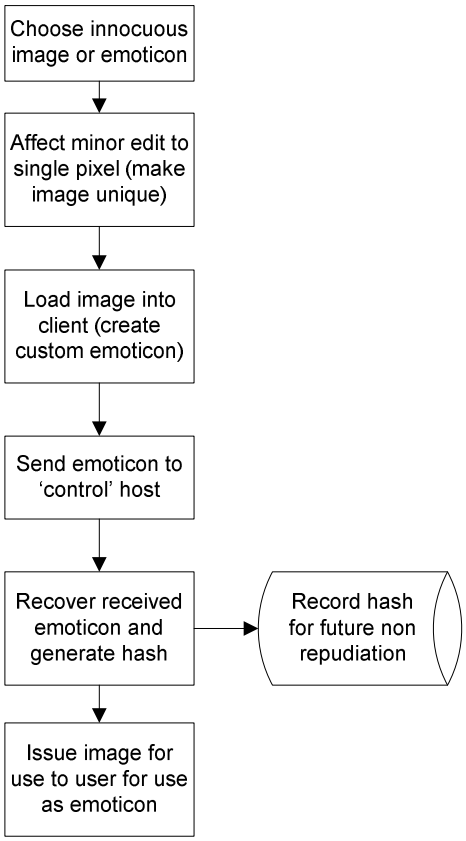


Figure 1 – Emoticon manufacture and issue

In the event a suspect's computer is seized and examined, and any emoticons are recovered these can be hashed and the hashes checked against the registry of stored emoticons. If a match is found this can be used in support of any legal action, see figure 2.

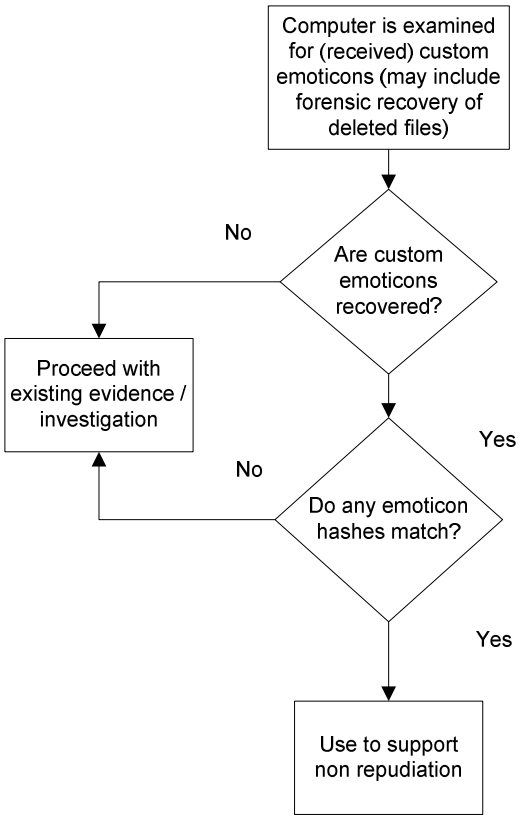


Figure 2 – Emoticon recovery

JUSTIFICATION

The process described here is part non-repudiation, part steganography and part illusion. Non-repudiation is the act of ensuring one party cannot dispute a contract, in this instance participating in an IM conversation. In the online world this is typically performed using digital signatures – each party signs the message with their private key, this can be verified (by other users) with their public key. This forms part of the Public Key Infrastructure (PKI). Digital signatures are viable only where both parties initially agree and sign with their private key. It is reasonable to assume groomers (and others of a criminal disposition) would be eager to disguise their identity from the start. This means any form of non-repudiation (in this case transmission of the emoticon) must be done in a way that does not require the cooperation of the receiver and does not attract their attention. This is sometimes referred to as a 'drive-by download'.

Stenganography, as discussed, hides information in plain sight. The minor change (pixel edit) performed to the emoticon changes the original in a way not readily detectable by the human eye but obvious when hashes are taken. Steganographing of the image is essential – it must be changed in order to be unique and the change must be hidden to avoid attracting attention. Electronic image steganography typically hides information using the least significant bits (LSB) of the pixel. In the system described here the content of the hidden information is not important, it serves as a marker which can (later) be used to identify the emoticon.

The process relies on the illusion that nothing special or unusual has occurred, the recipient has received a plain, unremarkable emoticon – something that will neither attract attention nor warrant further investigation. To be successful the emoticon will be received, ignored and remain 'unmolested' within the computer file structure.

If the emoticon were unusual the recipient might wish to investigate or re-used the emoticon. If it were suspicious the recipient might delete it[54].

The illusion is aided by the design characteristics of the protocol - unlike file transfer, emoticons and display pictures are handled automatically, they are sent transparently to the recipient (the recipient does not need to accept them). Further, with the *Microsoft* client the received emoticon is stored using a non-obvious filename buried deep within the file structure. Finding the file is a non trivial task.

The major governing factor is the existing technology. The technology was not designed to incorporate a non-repudiation function. This system has been conceived around these limitations.

A valid argument can be made that those that wish to harm children will merely purge their systems of the custom emoticon using a secure delete procedure such as those that conform to the Department of Defense DOD 5220.22-M standard[55]. As counter to this criminals know they can be traced by their finger prints – yet many crimes are still solved in this manner.

An argument can also be made that the custom emoticon can be faked or planted and as such fail to conclusively prove guilt. Finding the custom emoticon on a computer does not prove guilt or otherwise, it is circumstantial evidence that law enforcement organisations may use to build a case against a suspect. By itself it is conclusive of nothing, however since law enforcement would already have access to a suspect's computer in order to have found the custom emoticon it would serve to support their suspicions and corroborate their case. Comparisons can be drawn to the use of deoxyribonucleic acid (DNA), found at crime scene it does not prove a suspect is guilty of a crime but it does provide circumstantial evidence.

The custom emoticon may not be present on a suspect's computer, there are numerous reason why this may be so – the suspect may have used a different computer, found and securely deleted the emoticon, used a variant (third party) chat client that does not support custom emoticons, etc. As discussed previously the custom emoticon itself is conclusive of nothing, neither is its absence. It would support a legal action but would not be the sum of evidence.

The authors, while acknowledging these limitations, feel that any support to child protection/law enforcement activities is beneficial, as is any deterrent effect the approach may have on groomers.

CONCLUSION

This paper has looked at instant messaging and its popularity among children and youth. It has discussed technical measures used to protect children on the Internet and augmenting these with proper oversight and rules, and a good parent child relationship.

This paper has described the MSN Messenger Service Protocol and a method of leaving a 'digital fingerprint' on a recipient's computer during conversations. This method takes advantage of the (intended) operating characteristics of the protocol and does not involving 'hacking' or vulnerability exploitation.

The 'digital fingerprint' takes the form of a unique steganographed emoticon. It is made unique by making a minor change to the red, green or blue (and opacity, if supported)

[54] Deleted files can easily be retrieved using data recovery tools provided the data has not been overwritten.

[55] http://www.dtic.mil/whs/directives/corres/html/522022m.htm (accessed 12 September 2008).

channel. The change is undetectable to the human eye but is easily distinguished and thus identified by taking a hash of the result.

If a suspect's computer is found to contain customised emoticons these can be hashed and checked against a registry. If matches exist they can be used as circumstantial evidence in support of legal actions.

A unique steganographed emoticon could be included in a new third party MSN client, providing a simple, automated, 'turn-key' solution for the end user.

REFERENCES

Clayton, R. (2005). *Failures in a hybrid content blocking system*. Cambridge: University of Cambridge.

comScore. (2006). *Europe surpasses north america in instant messenger users, comScore study reveals.* Retrieved 25/02/2008, 2008, from http://www.comscore.com/press/release.asp?press=800

Crisp Thinking. (2008, 14 January). CRISP THINKING'S ANTI-GROOMING TECHNOLOGY ACHIEVES 98.4% ACCURACY IN INDEPENDENT TESTING [press release].

Hancock, J., Landrigan, C., & Silver, C. (2007). Expressing emotion in text-based communication. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, Calif. 929-932.

Kharif, O. (2 April 2001). BW online | april 23, 2001 | the man who brought a :-) to your screen. Retrieved from http://www.businessweek.com/bwdaily/dnflash/apr2001/nf20010423_785.htm

M2 Presswire. (2008). *Crisp thinking's anti-grooming technology achieves 98.4% accuracy in independent testing. | M2 presswire (january, 2008).* Retrieved 02/04/2008, 2008, from http://www.accessmylibrary.com/coms2/summary_0286-33698195_ITM

Microsoft, *Microsoft to publish MSN messenger service protocol to industry: Furthers commitment to instant messaging interoperability; provides solution to meet strong consumer demand for open communications.* (1999). Retrieved 23/09/2008, 2008, from http://www.microsoft.com/presspass/press/1999/aug99/protocolpr.mspx

Mitchell, K., Finkelhor, D., & Wolak, J. (2007). Youth internet users at risk for the most serious online sexual solicitations. *American Journal of Preventive Medicine, 32*(6), 532-537.

Mitchell, K., Finkelhor, D., & Wolak, J. (2001). Risk factors for and impact of online sexual solicitation of youth. *Journal of American Medical Association, 285*(23), 3011-3014.

Movva, R., & Lai, W. (1999). *Instant messaging and presence protocol.* Retrieved 23/09/2008, 2008, from http://www.hypothetic.org/docs/msn/ietf_draft.txt

Ozimek, J. (2008). *Porn, abuse, depravity - and how they plan to stop it • the register*. Retrieved 25/02/2009, 2009, from http://www.theregister.co.uk/2008/10/09/policing_internet_one/

Wolak, J., Finkelhor, D., Mitchell, K., & Ybarra, M. (2008). Online "predators" and their victims myth, realities and implications for prevention and treatment. *American Psychologist, 63*(2), 111-128.

5. THE ONLINE CHILD PROTECTION DOMAIN

In August 2002, two ten year old girls were murdered in Soham, Cambridgeshire (UK). Subsequent media coverage and government enquiries have elevated the subject of child protection to prominence in the public arena. New rules for those interacting with children could see 1-in-4 adults being vetted [103] as part of their work or social life.

During the same period worries about child safety on the Internet have risen. This is driven in part by alarmist television shows such as *To Catch a Predator*[56] but also by education and awareness campaigns pushed out by government and industry. This is often bundled with advice regarding Internet security and identity theft.

In today's society few crimes provoke more emotion and outrage than those committed against minors. Sexually based crimes against minors are considered particularly heinous; this is evident in both the media commentary and public reaction [1]. This public mood is undoubtedly shaped, in some part, by the public's perception of child sex offenders. The generally accepted depiction of these individuals is malevolent stalkers who prey on innocent children, lurking in the shadows patiently waiting.

Internet penetration in the UK is currently (September 2009) running at 76%[57] providing a new medium to those who prey on minors. Here too the public perception is of child molesters lurking in chat rooms and social networking forums popular with children to identify potential targets. Using information gleaned to manufacture online personae to mask their age and sexual intentions, enticing unknowing victims to offline meetings [2].

This has lead to much fear, uncertainty and doubt on the part of parents and guardians which may lead to the development of an overbearing parenting style which itself is having a negative effect on child development and societal practices [3].

[56] http://www.itv.com/News/tonight/episodes/Tocatchapredator/default.html & http://www.msnbc.msn.com/id/10912603 (accessed 15 Oct. 09).
[57] http://internetworldstats.com/europa.htm (accessed 12 Jan 10).

This chapter adopts a systems approach to understanding the online child protection ecosystem. It starts by defining what constitutes a paedophile and paedophilic actions and how this differs from those who have sexual interests in adolescences. The risks to children on the Internet from bullying, solicitation and unwanted exposure are enumerated. The Internet technologies and applications that are popular with children (chat rooms, gaming, social networks) are described and ranked in terms of the risk they pose. Existing technical measures (content filters, network level censorship) are described, where and when they are best used and their limitations are discussed.

5.1 Systems Engineering thinking applied to the problem

Systems Engineering is a branch of engineering that provides the theory and tools to manage the complexity inherent in modern engineering programmes, bridging the gap between the various technical specialties whilst maintaining oversight to eliminate programme drift and ensure customer requirements are met.

Systems Engineering encompasses technology, processes and management activities to define systems, capture and refine programme requirements, verify and validate design concepts, manage tradeoffs and configuration, develop architectures and integrate with existing technology and systems, including human systems (people).

The overarching goal is to deliver a high quality, fully functional product to the customer with minimal resources, waste and disruption while satisfying cost and time requirements. The complexity of modern engineering projects is beyond the skills and abilities of any individual or discipline. Systems Engineering seeks to understand and manage this complexity, bringing together integrated product teams from across engineering disciplines to approach the problem with a holistic view.

While the traditional tools, models and metrics of Systems Engineering are not appropriate to the problem of online child protection, the ethos of systems thinking has been applied to the problem space in order to understand factors including

- The motivation and requirement of stakeholders (groomers, children, parents, service providers, law enforcement, society, etc)

- The tradeoffs and conflicts between these requirements
- The threats children face on the Internet
- Threat vectors and technologies (chat rooms, instant messaging, email, webcams, social networking)
- The existing technological measures available to protect children and the limitations of these measures
- Procedural/Social techniques and behaviours that can be applied to protect children (parental rules, education)
- Online behaviour and other factors that increase or mitigate the risk children are exposed to
- The legacy system (the Internet)and how any solution must take into account it's design and limitation

Throughout this chapter the problem domain will be considered in a holistic manner that aims to derive the best use of resources to tackle child Internet safety, essentially providing the customer the best solution and best value for money.

5.2 What is paedophilia?

There is no standard agreed definition of paedophilia. The World Health Organisation (WHO) [4] defines paedophilia as a *sexual preference for children, boys or girls or both, usually of prepubertal or early pubertal age*.

The American Psychiatric Association in The Diagnostic and Statistical Manual of Mental Disorders (DSM)[58], specify the following diagnostic criteria for paedophilia [5]

A. *"Over a period of at least 6 months, recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving sexual activity with a prepubescent child or children (generally age 13 years or younger);*
B. *The person has acted on these sexual urges, or the sexual urges or fantasies cause marked distress or interpersonal difficulty;*

[58] Commentators have questioned the listing of paedophilia and other paraphilia (psychosexual disorders) in the DSM, suggesting they been listed without any scientific or rational basis [6].

C. *The person is at least age 16 years and at least 5 years older than the child or children in Criterion A."*

The generally acceptable definition of paedophilia among experts is a sexual deviation among adults for prepubescent children, where the adult is at least five years older than the minor. This is to exclude so called 'Romeo and Juliet romances' – relationships between similarly aged peers where one individual is under the age of consent. This definition is used in this thesis.

The terms hebephilia and ehebephilia describe adults with a sexual interest in pubescent and post-pubescent minors. Though often illegal and perhaps socially unacceptable, these are not typically considered sexually deviant as minors are sexually mature at these stages.

Kenneth Lanning [14], a veteran of the FBI Behavioural Science Unit describes a difference between paedophiles and child molesters; paedophiles have a sexual interest in children but may not act on these urges. A child molester engages in sexual activity with a child.

Lanning states that child molesters cannot be readily defined as being in one or more discreet categories, instead they fall along a continuum – see figure 12. At one end of the continuum are situational offenders, they do not have a preference for children and are more likely to be aroused by adult pornography. Frequently they will molest children they have access to (their own, children of friends, children in the neighbourhood) and are opportunistic. Often with low self-esteem and poor coping skills, they may molest under stress. Morally indiscriminate, they use and abuse all around – family, friends, co-workers, simply because they can. With feelings of inadequacy they see children as non-threatening sexual objects. They are likely to be less intelligent and of lower socioeconomic status.

At the other end of the continuum are preferential offenders - they prefer children. They molest children not out of insecurity but a sexual desire for children. They have a need for repeated sex with children and will have age and gender preferences. Introverted offenders often molest young children or strangers. Some may be sadistic

and need to inflict pain and humiliation. Diverse offenders will experiment with many paraphilia[59]. Preferential offenders are likely to be more intelligent and of a higher socioeconomic status.

Situation Offender (>more likely)	Preferential Offender (>more likely)
Impulsive	Compulsive
Lower Socioeconomic Status	Higher Socioeconomic Status
Less Intelligent	More Intelligent
MO (does what works)	Ritualistic
Opportunistic	Desire

Figure 12 - Sex offender motivation continuum (Lanning)

In the past the Internet offender was more likely to be of higher intelligence and status, but this continues to dissipate as Internet access becomes ever more ubiquitous.

When discussing technical measures this thesis uses the term groomers to describe those using the Internet to contact minors for sexual means who may or may not continue to molest children. *The author feels the term predator too emotive a term for scientific discourse.* When discussing sociologically and non technical measures, or where it is necessary to distinguish, protection measures will be defined in terms of pre- or post –pubescent children. The term child or minor is used to describe an individual under the age of legal consent in the UK (16). Some of the research referenced in this thesis has been conducted in jurisdictions outside the UK (primarily the United States) where the age of consent differs. Where possible this is noted and accounted for in the analysis.

[59] Paraphilia are psychosexual disorders including (but not limited to) voyeurism (watching others have sex), urophilia (involving urine, colloquially 'golden showers'), scatolgia (obscene talk). Consensual adult activities including role play and superficial, sexual fetishism (e.g. use of sex toys) does not necessarily constitute paraphilia.

5.3 Risks to children online

Although empirical data exists on solicitation, bullying and exposure to inappropriate content it is difficult to define metrics for how harmful this may be. Research indicates that a minority of youth [43] [71] are troubled by online solicitation or inappropriate content, most simply ignore it and consider it 'the price of doing business on the Internet'.

Side-by-side comparisons of surveys and other literature is extremely difficult, there is no agreed definition of solicitation or bullying. Again the *Crimes Against Children Research Center* (in the YISS surveys [72] [73]) have a concise definitions and these are clearly described in their papers. This lack of agreed definition and standards means researchers can have wildly differing findings in the same area, with same papers seeming to contradict others, for example

- Research suggests that around half [66] are not bothered by pornography and 25% (of children unintentionally exposed to pornography) are extremely distressed and 20% are mildly distressed [64]. *There is no indication as to what form the pornography takes (topless soft-core, BDSM[60] hardcore).*

- Exposure to pornography leads to deviant sexual practices and behaviour [74], or pornographic exposure leads to no increase and perhaps even a decrease in sex crimes [32]and little scientific evidence about the risk to children [64] [66].

The difficulties involved in researching this area are further explored in subsequent sections.

Vulnerable children – those with a history of abuse (physical, psychological, sexual), children who have a poor relationship with their parents, those who are isolated from their peer group, lacking in self confidence, with low self-esteem or who are bullied are at greater risk of all types of Internet victimisation [19] [42] [66] [71].

[60] Bondage, Discipline, Sadism and Masochism (BDSM), is form of sexual role play and expression based are power and pain.

5.3.1 Bullying

Grooming, sexual solicitation and inappropriate content receive most attention when online child protection is discussed. Bullying however is more than twice as common as solicitation [41]. Figures also show children who are victims of bullying are more likely to report being distressed or frightened (33%) [42] than those who are targets of online solicitation (25%) [43] [44].

Willard [45] enumerates the various forms of bullying as flaming (online fighting), harassment (insulting messages), denigration (rumours and gossip), impersonation, outing (revealing secrets), trickery, exclusions and cyberstalking (intense and repeated harassment and denigration).

Ybarra and Mitchell [42] estimate 30% of youth are involved in bullying as either victim, bully, or both. They also found that blocking software and household rules on Internet use had no significant effect on bullying. Messages about Internet behaviour or reducing time spent online were insufficient in tackling the problem. They recommend promoting positive parenting and youth involvement in anti-bullying initiatives.

The definition of school bullying typically requires 3 criteria

- Aggressive acts, verbal included with intent to harm
- Repetition
- Imbalance of power.

Because of the widely varying characteristics of cyber bullying, Wolak *et al* [56], encourage use of the term online harassment instead. Their research suggests 9% of children are harassed online, 55% of these harassments were from people the victim had not met in real life. Whether these harassers were known in real life but adopted pseudonyms (sometimes called sock puppets) is not known. Also interesting among their findings was the difference between harassment from known real world peers and online only contacts. Further they found that half (49%) of targets terminated their bullying by blocking or ignoring the harasser.

Patchin and Hinduja [69] champion the term cyberbullying and equate it more to violence than harassment - it can lead to violence, injury and even death for both bully and victim and is commonly a driver for aggression, depression and abuse. They describe how the psychological abuse of cyberbullying may be worse then physical abuse – 31.9% of victims reported being affected at school, 26.5% affected at home, 20.4% affected with friends. These effects, occurring during such formative periods of a child's life may have consequences stretching well into adulthood.

Campbell [75] too feels that cyberbullying may be more severe than face-to-face bullying. She cautions against simply punishing bullies, as this is of questionable efficacy in the offline world and would be impossible to apply against anonymous cyberbullies. She encourages peer group involvement i.e. bystander intervention, this may be effective in public forums such as social network profiles and chat rooms, but would not be applicable in private environments such as email or text message.

In order for there to be a proper understanding of the problem and meaningful comparison of literature, researchers need to develop a common language of agreed terms and definitions. In addition there needs to be qualitative research conducted to distinguish relatively minor incidents from serious events and in the harm done by bullying. Continuing research is also required to understand the long term effects of online bullying.

5.3.2 Unwanted Exposure

In the UK there are numerous laws that restrict access to offline media content. The *Obscene Publications Act 1959* outlaws materials (texts, images, audio and video) that "tend to deprave and corrupt" [60]. The *British Board of Film Classification (BBFC)* is an NGO (funded by the film industry) to classify videos[61]and some computer games (those with violence, sexual content, or encouragement of illegal activities e.g. *Grand Theft Auto*), under the *Video Recordings Act 1984 / 2010* [61]. Sections 63-66 of the *Criminal Justice and Immigration Act 2008* prohibits

[61] Local authorities have the power to define the circumstances under which films are shown in cinemas, but they rarely deviate from BBFC advice.

pornography that realistically depicts threats to a persons life, serious injury to breasts, the anus or genitals, sex with a corpse, or sex with animals [63].

The UK has not seen a successful prosecution under the *Obscene Publication Act* in over 30 years, indeed a trial collapsed against a UK civil servant in 2009 when the prosecution offered no evidence [62]. Selling any video without BBFC certificate (or exemption) on the packaging is an offence.

These acts ensure physical media available in the UK adhere to certain standards and to prevent minors acquiring age restricted media. The Internet allows individuals to distribute content without the costs associated with physical distribution and without age verification checks. The structure of the Internet allows individuals to legitimately post/host content that would be illegal under UK law.

Responsible adult industry websites have warnings on their landing pages and content labelling indicating they are for adults only [76]. Disreputable Internet distributors attempt to maximise pornographic exposure to all (often in conjunction with other illegal activities e.g. SPAM, malware, viruses).

In addition to pornography, parents may be keen to protect children from hate sites, bigotry, violence and sites promoting or glamorising illicit activities such as drug taking.

The generally accepted position is that exposure to pornography is distressing to children and may shape their long term sexual development. While some may find it distressing or embarrassing, there is no significant evidence of long term harm [64] [65]. Younger children lack the maturity to understand pornography; they are usually annoyed or disinterested by it. Older children tend to find unsolicited pornography an irritant like spam. Some minors may be distressed by extreme pornography – explicit scenes of bondage where pain and physical trauma are involved.

The limitations of this evidence should be noted - the effects of pornography on minors cannot easily be researched because of ethical issues, the body of work is small and no consensus exists (indeed some findings are contradictory) [66].

Exposure to hate content can be a problem as children often lack the critical skills necessary to evaluate source material. There is no scientific consensus on the effects of exposure to violent content [66].

5.3.3 Solicitation

'Stranger danger', 'Online groomer', 'Internet predator' – these terms are frequently bandied about, particularly by the more shrill and alarmist elements of the media [2] [14].

Undoubtedly molestation is a horrific crime and a parent's nightmare, but how does 'real world' solicitation compare with anecdotes from the media? Online groomers are typically described as combing the Internet looking for children, befriending them with fake personae, tricking them into performing online sex acts and arranging real life meetings [67]. While there are undoubtedly individuals who operate in this way, their behaviour is atypical and they represent a very small minority of groomers. The true picture of solicitation and grooming is someway removed from this depiction.

Research shows the typical groomer does not conceal his[62] age or intents, groomers may shave a few years off their true age but by and large they are honest about their age and bring their sexual desires into the conversation early on [2]. The overwhelming majority of Internet initiated sex encounters (with minors) involve post-pubescent minors. Typically such individuals are physically mature and sexually aware, they are willing, in some cases enthusiastic participants in their own abuse. Cases involving prepubescent minors are virtually unheard of and it may be the case that the child has already been sexualised – groomed or abused within the friends and family circle. *Internet initiated grooming of prepubescent children is so rare because of the difficulty engaging them sexually (something beyond their maturity and understanding) and because they are more closely supervised online.*

Society refuses to accept the image of minors as active agents; happy to meet older men for sex, yet these headstrong adolescents typically describe a strong romantic link or report being in love. If/when these relationships are uncovered nearly half of victims refuse to cooperate with law enforcement against their abuser [14]. Society's

[62] The overwhelming majority though not all Internet sex crimes are committed by men.

refusal to accept the evidence, instead falling back on an artificial reality - helped by egregious media accounts and in some case official guidance and literature, could be increasing the risk to children. If children and parents are misinformed of the risks, how can they adequately protect themselves and their children?

As with elsewhere in the child protection ecosystem, terms – in this case solicitation - are frequently bandied around without explanation or context. The literature often reports 1-in-7 youths are solicited by online predators, citing research from the University of New Hampshire. The research authors concerned about the misuse of this statistic put out a release explaining it [68]. The solicitations did not necessarily come from predators – many came from the youth's peer group, often solicitations were limited to vulgar comments like, "What's your bra size?". Two thirds of the victims were untroubled by the solicitations and almost all handled the situation easily and effectively.

While still a source of concern, online solicitation is not as pervasive and filled with dishonest predatory men as many have been led to believe.

5.4 Threat vectors

To ensure the reader is fully informed this section will briefly look at the technologies the Internet provides children and the risks they pose.

Chat rooms are primarily a web based platform where individuals can chat in a public environment or select users (from those logged into the site) and engage in private communication. It is analogous to a party, conversations can be public or 'taken quietly' to the periphery (private one to one). Until recently chat rooms posed the greatest risk of solicitation to minors [43], IM now poses the greatest risk (see table 1). Chat rooms involved in Internet initiated sex crimes are typically keyed towards teens, geographical locations, dating and romance and in some potentially worrying cases, sexual encounters between adults and children [17].

Instant Messaging (IM) provides a private medium for users to chat one-to-one, or in groups. IM requires users to have an existing relationship. IM has support for presence (reveals whether the user is logged on / away / busy, etc) and many IM platforms also support audio and video communications. Bullies and solicitors may be

finding children's IM detailed from sources such as blogs or social networking profiles. IM clients typically allow users to block unwelcome contacts, indeed this is what the majority of children do [68].

Online gaming provides a multiplayer element to traditional computer games. Massive Multi Player Online Games (MMPOG) further extend the concept with hundreds, even thousands players involved in the game.

Internet Relay Chat (IRC) is similar conceptually to chat rooms – themed 'rooms' called channels, support group discussion and one-to-one communication via private message.

Weblogs (blogs) are a form of online public diary, in addition people may leave comments. There is concern that personal information written in a blog could be used to build a profile of the child which could be used to harass or solicit, however this does not appear to be happening [2] [41].

Short Message Service (SMS) text messages – while not strictly an Internet application it is included here for completeness. Messages up to 160 characters in length are sent between mobile telephones or often between the Internet and a mobile phone. Mobile phones are a source of concern because of their pervasiveness, as communications become more unified and 'smart phones' with more features become the norm, traditional Internet threats will continue to migrate to this platform.

Social Networking has become such a major component of the Internet in recent years it is covered in more detail in the subsequent section.

Many of these technologies are converging, for example the social networking platform *Facebook* includes a chat client, essentially IM functionality that can be used independently of the main site. Many individuals also use their SNS to blog, though to a lesser extent than those who maintain a full blog.

The rapid evolution of the Internet and speed with which services take hold or fall from public favour means usage and risk profiles can quickly become outdated.

Table 1 shows a breakdown of solicitation and harassment by platform/application, based on responses from (a statistically significant sample of) 10- to 15-year-olds.

	Solicitation[63], %	Harassment/Cyberbully , %
IM	6.2	16.5
Chat rooms	4.5	5.9
SNS	4.2	8.2
Email	2.4	6.2
Games	2.8	7.8
Blogging	0.8	2.5

Table 2 - Minors solicited or harassed online in the past 12 months, (Source [41]).

One activity known to increase the risk to children is discussing sex on the Internet with strangers [77]. However Subrahmanyam *et al* [78] found that teens' primary source of sexual information is their peer group (little is known of the substance or quality of this information). Adolescents discuss sex as way of understanding and controlling their emotions. Maintaining open communications can serve as a coping mechanism for sexual expression, particularly when providing an anonymous forum for discussing embarrassing issues. They suggest it might be preferable to have children engage in online relationships and cybersex[64] as this entails less risky behaviour, particularly unprotected sex. Cybersex and online relationships may allow girls (the greater risk group) to exercise more control as they are recognised to possess superior communication skills.

5.5 Social Networking

Social Networking has seen a meteoric rise in recent years, in January 2009 the two top Social Networking Sites (SNS) - *Facebook* and *MySpace* recorded over 2 billion visits [15]. The format of SNS differs between providers but typically a user would post information about themselves, photos and links to friends' profiles. Some SNS

[63] Defined as all unwanted online requests to youth to talk about sex, answer personal questions about sex or do something sexual. Most were limited to brief online comments or questions in chatrooms or instant messages. Many were simply rude, vulgar comments like, "What's your bra size?"

[64] Cybersex is a form of computer mediated communication where users send each other sexual explicit messages – it may include fantasy, role playingand real life masturbation.

profiles have privacy controls limiting the amount of information that can be seen by strangers, others show everything.

Does having a SNS profile increase the risk to minors online? Smith at Pew Internet [16] found that 49% of teens use social networks to make new friends and 32% were contacted by a stranger[65]. 7% of teens felt scared or uncomfortable as a result of contact with a stranger, this was far more likely among girls. No association was found between the type of information (name, address, email, school, etc) posted and the prevalence of stranger contact, but those who post photos online or had profiles were more likely to be contacted by a stranger.

Most sources [18] contend that exposing personal information on the Internet is a risky behaviour. Whilst possibly counterintuitive, research indicates [2] [41] that posting personal information on a SNS or blog does not increase the likelihood of harassment or sexual victimisation. *However limiting the amount of personal information openly available on the Internet can be valuable in keeping such information private from college admissions tutors and future employers* [17]. *It may also be valuable to minimise such disclosures as a defence against identity theft.*

Despite the 'apparent' increase in risk that accompanies SNS use – exposure of personal information, posting photographs and the desire to meet new people, social networking does not increase the risk to minors [19]. In fact youths report considerably less instances of both sexual solicitation and harassment in social networking, than in other Internet activities. A survey of 1,588 10-15 year olds revealed 15% reported an unwanted sexual solicitation on the internet in the previous 12 months; only 4% reported a solicitation on a SNS specifically. 43% reported being the target of harassment on the Internet; only 9% reported harassment on a SNS. Among victims more solicitation and harassment took place on IM (43% and 55% respectively) than social networking (27% and 28%) [41].

Social Networking has been a hot topic for legislators. In summer 2006 the US Congress held four hearings on social networking, resulting in calls for greater

[65] Defined as someone with no connection to you or any of your friends

regulation and oversight of these sites with such claims as "[social networking is a] virtual hunting ground for predators" [41].

In the period 2007-2008 the SNS *MySpace* removed the account of 90,000 registered sex offenders [20]. Roy Cooper, the North Carolina Attorney General[66] (AG) said "Predators are going to trawl in these areas where they know children are going to be. That's why these social networking sites have the responsibility to make their sites safe for children." boyd [sic] [21] accuses these AG of political grandstanding and fear mongering, pointing out that many offences (including public urination [22]) can lead to sex offender registration, none of which involve children. Nor is it illegal for sex offenders to have social network profiles (unless this is a condition of their licence (probation)).

The Internet Safety Technical Task Force (ISTTF) [19] was created in 2008 by the *US Attorneys General Multi-State Working Group on Social Networking* to assess the risks to children posed by such technologies.

Pennsylvania AG Cooper uses the *MySpace* figure to attack the findings of the ISTTF as being "outdated and inadequate". Willard [23] analysed the work of the Pennsylvania *Child Predator Unit*[67] arrest data, based on press releases from the AG. She found 143 articles containing the word 'predator' over a four year period. Her finding revealed

- 8 incidents where the Internet was used to form the relationship (4 were reported by parents or teens, 4 were discovered from files after sting operations), 5 of these led to sexual encounters.
- 166 arrests are reported as the result of sting operations (144 were in chat rooms, 11 IM, 9 unspecified).
- Only 12 predators lied about their age.
- No arrest took place as a result of SNS communications.
- In one instance involving a teen victim communications took place on MySpace, this was a re-arrest.

[66] The position of a state Attorney General is political, the majority of Attorneys General are elected included all mentioned here.
[67] Formed by AG Cooper in 2005 using specially trained agents and prosecutors [70]

- One was a police officer who sexually abused girls he met in the line of duty. He had a *MySpace* account linked to a teen girl – there is no assertion that this led to sex.
- In one sting the predator offered a *Facebook* link.
- The sting profiles set up by the Child Predator Unit did not use *MySpace* protective features.
- There appears to have been no *MySpace* stings in the past two years despite honey trap profiles

AG Corbett stated the Internet is the primary means for predators contacting victims. In the past 4 years the Child Predator Unit (10 staff) have discovered 8 incidents of sexual abuse involving actual teens where the Internet was used to form a relationship. During a single year *Pennsylvania Coalition Against Rape*[68] recorded 9,934 child victims of sexual abuse.

The *Pennsylvania Coalition Against Rape* reported seeing increased use of digital media to manufacture child porn by family and friends, not SNS based offences [23]. *The author suspects the ubiquity of digital media especially digital cameras (devoid of need of third party processing) would see an increase in all photography of an illicit nature (e.g. images of drug taking/street fighting).*

In one press release the sting took place in an 'Internet romance chat room' [23]. *This suggests that while the victim may have been exploited, they would probably not have been misled (at least about the 'predator's' intention).*

These results are inline with the findings of the ISSTF (and others – see table 1), chat rooms and IM pose a far greater risk to children than SNS and the overwhelming majority of child abuse is at the hands of family, friends and individuals known to the family, not Internet strangers [14] [19].

Willard is also scathing of age verification for accessing SNS and other services [24] she argues that any non-industry wide system would fail as users would go elsewhere,

[68] Founded in 1975, it operates a network of rape crisis centres and engages in advocacy work http://www.pcar.org/about-pcar (accessed 01 June 2010).

whilst wide scale systems would invite [civil liberties] protests and evasion technologies. Evasion technologies likely means technologies similar to Internet 'anonymisers' [sic] and Digital Right Management (DRM) strippers used to remove copyright protection from digital media.

The *Child Exploitation and Online Protection* (*CEOP*) Centre has been demanding sites display its 'Panic Button' (a large graphic, linking to the *CEOP* website) on user profiles. The *CEOP* button is displayed on the *Bebo*[69] SNS and the *Window Live Messenger* (*MSN*) chat client [57]. Despite the relative safety of SNS, *CEOP* CEO Jim Gamble, with the support of UK Home Secretary Alan Johnson, has been publicly attacking *Facebook*[70] for refusing to publish his organisation's button. [58]. *Facebook* has dedicated safety teams providing 24 hour support in 70 languages [57]. Forum commentators have also noted that *Facebook* is an international site and *CEOP* is a UK organisation. BBC online forums [59], generally considered to be a barometer of public opinion in the UK, show little support (10% of posting as of April 2010 – six months after the article) for *CEOP*'s position.

This paints some politicians as either woefully misinformed or exploiting a public fear for political reasons. Regardless of the motivation this has a number of undesirable consequences. It leads parents and the public to false conclusions, leaving them poorly informed about the risks, this in turn leads them to make poor decisions regarding children in their care. Further, it may result in unnecessary and largely ineffective legislation that does little to protect children on the Internet. This in turn may have the knock-on effect of giving parents and others a false sense of security.

Dangers to children do exist on the Internet, but social networking is not the greatest threat. Though probably not a good idea, posting personal information on the Internet on blogs and SNS pages does not increase the risk to children.

[69] http://www.bebo.com (accessed 6 April 2010)
[70] http://www.facebook.com (accessed 6 April 2010)

5.6 Child Pornography

The COmbating Paedophile Information Networks in Europe (COPINE) scale provides a typology of Internet child pornography images (also called child abuse images). In the UK the *Sentencing Advisory Panel*[71] (2002), advising on cases (in England and Wales) involving child pornography adapted the scale – dropping levels 1-3 on the grounds that nakedness is not indicative of indecency and combined levels 4-6 [51]. The COPINE Scale is increasingly being used as measure of the seriousness of an offence and even how dangerous an offender might be. See table 2 for details.

CS	SAP	Description
1	N/A	Indicative: Non-erotic and non-sexualised pictures showing children in their underwear, swimming costumes etc. from either commercial sources or family albums. Pictures of children playing in normal settings, in which the context or organisation of pictures by the collector indicates inappropriateness.
2	N/A	Nudist: Pictures of naked or semi-naked children in appropriate nudist settings and from legitimate sources.
3	N/A	Erotica: Surreptitiously taken photographs of children in play areas or other safe environments showing either underwear or varying degrees of nakedness.
4	1	Posing: Deliberately posed pictures of children fully clothed, partially clothed or naked (where the amount, context and organisation suggests sexual interest).
5	1	Erotic Posing: Deliberately posed pictures of fully, partially clothed or naked children in sexualised or provocative poses.
6	1	Explicit Erotic Posing: Pictures emphasising genital areas, where the child is either naked, partially clothed or fully clothed.
7	2	Explicit Sexual Activity: Pictures that depict touching, mutual and self-masturbation, oral sex and intercourse by a child, not involving an adult.
8	3	Assault: Pictures of children being subject to a sexual assault, involving digital touching, involving an adult.

[71] http://www.sentencingcouncil.org.uk/ (accessed 29 July 2010)

9	4	Gross Assault: Grossly obscene pictures of sexual assault, involving penetrative sex, masturbation or oral sex, involving an adult.
10	5	Sadistic/Bestiality: a) Pictures showing a child being tied, bound, beaten, whipped or otherwise subject to something that implies pain. b) Pictures where an animal is involved in some form of sexual behaviour with a child.

Table 3 - COPINE Scale (CS) and SAP = Sentencing Advisory Panel (SAP)

Those who have been subjects in child abuse imagery frequently report feelings of guilt, depression, aggression and difficulty concentrating [14] [51]. These feelings may be enforced by knowledge that such imagery continues to exist.

It is difficult to imagine a reasonable person who would condone child pornography (when considering the suffering of its subjects), however Computer Generated Images (CGI), sometimes referred to as 'Photoshopping' can create photo realistic images where no child is ever harmed. CGI and other artificial images can show children in a sexualised fashion without any victim. For example *lolicon* and *shotacon* - sub categories of manga and anime (Japanese comics) features sexualised imagery of prepubescent girls and boys.

The remainder of this section, when referring to child abuse imagery and child pornography shall mean material that has been created where no child was harmed, unless explicitly stated otherwise.

With no victim, where no child is harmed, the only rational argument against child pornography is that it leads to a normalisation of inappropriate feelings, which in turn leads to an increased risk of (real life) child abuse. This is the generally accepted view [46] [48] [79].

Research indicates that the spread of wholesome adult pornography (featuring no children) and easy access to it, especially via the Internet has led to a reduction (or at least no increase) in rape, sexual assault and sexual abuse [31] [32]. Pornography allows the user to satisfy their fantasies and provides a cathartic release; this can be as

part of a multi-person sexual relationship or individually in masturbation. If pornography leads to a reduction in the rate of rape and sexual assault, why does exposure to child pornography lead to an increase in the rate of child abuse? Are seemingly identical activities (viewing pornography) leading to diametrically opposing outcomes?

The fear that artificial child pornography (real child pornography was already prohibited under the *Protection of Children Act*[72]) triggers an inevitable spiral into child abuse has led (in the UK) to section 49 of the *Coroners and Justice Bill*[73] – *'Possession of prohibited images of children'*. This outlaws child abuse images that do not contain real children (computer generated imagery, cartoons, drawings). During the consultation period the Home Office and the Police expressed disgust at some cartoons and manga, but no evidence was found that such images could fuel abuse by reinforcing potential abusers' inappropriate feelings towards children [46].

The consultation document cites a single case in which police raided an individual and found them in possession of only cartoon images. They had no choice but to hand the material back and allow its owner to walk free. "It is possible that this exemplifies that category mentioned by Lord Hunt in summing up on extreme porn: people whom the police would like to "do something about", but who haven't actually broken any laws" [46].

A well known Internet image series showing members of the *Simpsons* cartoon family in lewd poses would fit into this category [47], indeed Australia has seen two convictions (2008 & 2010) for possession of this material [52][53]. Technically the Olympics 2012 logo also falls foul of this legislation (some reporting it resembles cartoon character Lisa Simpson performing fellatio) [50], *however its establishment backing makes any investigation unlikely*.

Liberal Democrat MP Jenny Willott is among those who question the veracity of claims regarding the viewing of child abuse imagery leading to real world abuse.

[72] http://www.statutelaw.gov.uk/content.aspx?parentActiveTextDocId=1502057&ActiveTextDocId=1502059 (accessed on 01 June 2010)
[73] http://www.publications.parliament.uk/pa/cm200809/cmbills/072/09072.25-31.html#j3_100a (accessed on 01 June 2010)

Raising the issue in a parliamentary cross-party committee hearing of the *Coroners and Justice Bill* [48], she later elaborated

"This is the case with the sections on possession of Cartoon images... We are being asked to choose between two conflicting world views: on the one hand there is a belief in the 'slippery slope', that looking at images habituates individuals to the actions involved and can increase the risk to children; on the other it is argued that these images act as a release and actually reduce the incidence of harm.

"It is worrying that we appear to be legislating on this subject without hard evidence either way – especially when getting it wrong could have such serious implications for children. We have passed laws against possessing indecent images of children, because such images are evidence of harm committed: that is clearly not the case with CGI imagery and before we criminalise it, we should be prepared to come up with evidence of harm caused by the impact of seeing that image." [49]

There is no consensus about the potentially 'enabling' qualities of child pornography, some experts argue it is a stepping stone to abuse [79], others contending that it helps prevent child abuse [80]. There is no significant evidence of a causative link [2].

If child pornography, where no real child is harmed, does increase the risk of real world child abuse, then Section 49 of the *Coroners and Justice Bill* is sound legislation. However if the bill was crafted, based not on evidence, but good intentions, subjective morality, or a desire to 'appease decency' it maybe unnecessary or even increasing the risk to children. Rather than being a driver for child abuse such images might serve as substitute for it. It is alarming to consider that laws designed to protect children may in fact be endangering them.

Whether such images increase or decrease the overall risk to children is beyond the scope of this thesis, it is certainly an interesting and valuable area for psychological research. What is of concern is that without clear evidence, good intentions and knee-jerk responses might be worsening the situation. The author calls for more research and evidence based legislation. *Conducting research on the matter would probably be illegal in the UK.*

5.7 Current safety / prevention message

At a 'grass roots' level UK government agency *CEOP* has an excellent website promoting Internet safety aimed at children. *Thinkuknow*[74] [sic] is divided into three age categories (5-7, 8-10and 11-16) with advice relevant to the target age group. It introduces children to the idea that things may not be as they appear, that individuals may be deceptive and concepts such as limiting the amount of personal information disclosure. It is not without short comings; in a section on IM it says "*Be careful not to share too much information with other people in chat rooms. You don't know who could be listening in or what they might do with that information.*" [81]. This is (in the author's opinion) probably good advice, but it is known that youth perceive this as unrealistic and will tend to ignore it [43]. In a section on social networking it says "*Be careful who you agree to accept into your forums / private chat areas.*" [82] yet adolescents often use the Internet to make new friends [16] [83].

The 'ClickCEOP' button is a graphic linking to the *CEOP* website where individuals can report abuse or seek advice [85]. In the media *CEOP* have described the button as analogous to a burglar alarm that shows (users) children are in a 'protected' environment [59]. *CEOP* have even linked the death of Ashleigh Hall (the 17 year old who met her killer on *Facebook*[75], which does not have the button), to the absence of the button [84]. Home Secretary Alan Johnson and Liberal Democrat MP Chris Huhne joined the debate attacking *Facebook* for not featuring the button [86].

Though not perfect, *Thinkuknow* offers useful, practical advice and age relevant guidance to children (also parents and teachers). At the same time *CEOP* are 'muddying the waters' in the media – the *ClickCEOP* button *may* act as a deterrent but it provides no protection, it doesn't monitor activities, profile behaviour or respond to events. Ashleigh Hall also used *MSN* to communicate with her killer – MSN does feature the *ClickCEOP* button (she did not use *CEOP* resources but she would probably have been aware of them). This media campaign and the supporting

[74] http://www.thinkuknow.co.uk/ (accessed on 29 July 2010)
[75] *Facebook* have since reached an agreement with *CEOP* whereby users can opt to have the *ClickCEOP* button on their profile – *CEOP* wanted it to appear on every profile.

comments from government ministers and parliamentarians serves to confuse parents, children and the public with spurious and unnecessary information.

5.8 Examination of current protection software

This section will examine some of the existing technical measures available to protect children on the Internet discussing both their strengths and limitations.

5.8.1 Content Filtering

Software installed on the home computer can restrict access to certain websites, for example those with pornographic or violent content. Typical users can block content based on categories such as sex, criminal skills, drugs etc. Entire applications and protocols may also be blocked but this is typically permit/deny rather than content filtering.

Content on the Internet can be identified through three ways, the owner can mark the content, a reviewer can categorise it or automated processes can categorise it. All three methods involve some measure of subjectivity (the automated process has to be programmed/developed by a person). Unlike responsible pornographic webmasters, Internet criminals often host malware under the guises of pornography, such individuals are unlikely to employ content marking techniques, as their goal is to maximise infection. Manual review of sites is un-scalable given the enormous amount of content on the internet – the claims of content filtering providers have been challenged before the US Congress [25]. Automated content review analyses content on the fly, traditionally this has been performed with keyword spotting, this approach has a very high failure rate [25] [19] [64].

There is evidence that the US conservative Christian right is influencing the quality of content filtering software, introducing a bias against non-traditional religions (new age spirituality) and sites providing sexual health information to homosexuals [87].

Content filtering is beneficial for preventing accidental exposure to unwelcome material and is suitable for protecting younger naïve children. Against head strong teenagers, especially technically competent individuals seeking such content, filters offer minimal value.

5.8.2 Monitoring

It is possible for parents to monitor and review their children's' Internet activities, this might include examining website histories, reading emails, IM logs, system logs, even installing key loggers. Monitoring can be performed automatically and holistically or manually on a per event/application basis. It can be conducted covertly or with the child's knowledge.

If a child uses the Internet frequently or for extended periods, or in houses with more than one child using the Internet, monitoring quickly becomes unscaleable. Older children and teens may object on grounds of trust and privacy. Research indicates that parents struggle to decipher youth lingo on the Internet [25] which limits the efficacy of monitoring and as with filtering, technically competent youths will be able to circumvent this or use an Internet connection elsewhere (school, library, Internet café).

Monitoring may encourage good behaviour on the part of youths who know they are being monitored and may be useful to monitor children during periods of concern, e.g. a parent worried their child is taking drugs or involved with the 'wrong crowd'.

Monitoring does not correlate to reduced risk of Internet victimisation [17], but is indicative of lower likelihood of stranger contact [16].

Martin, Durbin, Pawlewski and Parish [90] present a novel approach for monitoring/review in Internet 'chat' environments. Intelligent text analysis is used to identify potential 'conversations of concern'; these conversations would then be anonymously reviewed by other parents in a community effort. Parents would, if necessary, receive feedback alerting them that their child was engaged in potentially unsafe behaviour, but not the original 'message of concern'. In this approach children would be allowed privacy (from their parents) but would be monitored by technology and community members. In return for this service parents would have to anonymously evaluate 'conversations of concern' from other children.

Crisp Thinking[76] has developed a tool described as an *Anti-Grooming Engine (AGE)* which builds a behavioural profile of the child's chat using Bayesian inference [91]. *Crisp Thinking* claim 98.4% accuracy in independent tests [92] but beyond this headline figure, results, test data and methodology are not in the public domain[77]. While the range of products offered by *Crisp Thinking* are probably useful for helping to protect younger children, their value in protecting teenagers who are discussing sex within their peer group are more questionable. While the algorithms and test methodology remain secret, the efficacy must be treated with scepticism. *Crisp Thinking* has applications that can be applied in the network at ISP or site level, thereby preventing local bypass measures.

5.8.3 Moderation and Pre-moderation and Wall Gardens

Many message boards, forums and sites that allow user generated content and feedback moderate content. This typically takes the form of site administrators and moderators who review and remove content that violate site rules e.g. obscene content, spam, trolling (deliberately inflammatory, often off topic content, designed to offend).

Moderation naturally has a time delay (until the moderator review the content), is subjective (based on the moderators bias) and is clearly not scalable. Moderation is useful for enforcing good behaviour and etiquette.

Pre-moderation sees all user generated content and feedback screened by moderators before it is publicly viewable. This process is used in forums targeted at young children; the BBC[78] employs trained individuals to screen messages for signs of bullying, harassment and other inappropriate content. Pre-moderation is extremely resource intensive and only suitable in very limited circumstances.

Walled Gardens are environments where only 'whitelisted' content is permitted – all content is banned until it is has been evaluated and approved. *KidZui*[79] is an example

[76] http://www.crispthinking.com (accessed 10 Nov 2009)
[77] The authors have asked to see information
[78] http://www.bbc.co.uk/chatguide/glossary/moderation.shtml (accessed 03 August 2009)
[79] http://www.kidzui.com/about_us (access 03 August 2010)

of this technology but it is extremely limiting in what in can offer and is only suitable for very young children using the Internet with minimal parental oversight.

5.8.4 Network Level Censorship - IWF and CleanFeed

In 1996 driven by public concern about paedophilic content on the Internet the ISP association (ISPA) created the Internet Watch Foundation (IWF)[80] [81].

The *IWF* maintains a list of child abuse websites; the list is updated twice daily and contains 800 – 1,200 live sites. About 50 URLs are added daily [8].

CleanFeed is an ISP level filter used by major ISPs. *CleanFeed*[82] is a two stage filtering process – if a request is made for information hosted at a 'suspect IP address' (an IP known to host child abuse images) it is routed to a HTTP proxy. Otherwise traffic is sent to destination unmolested. Suspect traffic is examined in detail at the proxy and compared to the *IWF* blacklist [9], requests for child abuse content is 'blackholed', requests for legitimate content at 'suspect IP addresses' are permitted. *CleanFeed* can be bypassed using international proxies.

In 2009 the *IWF* received 38,173 reports, it took action against 8,844 pages across 1,316 sites. *IWF* notes that despite rapid expansion of the Internet in recent years, the volume of commercial child pornography has remained static [93]. *There is no breakdown of these figures into genuine vs computer generated imagery; this might include computer generated images on servers in the US which enjoy first amendment protection.*

ISP *Zen* (and others) have declined to implement the *IWF* blacklist recommended for blocking child sex abuse images [7]. Zen cited questionable efficacy of the system. The *Children's Charities' Coalition on Internet Safety*[83] (CHIS) has demanded government action saying self regulation has failed.

[80] http://www.iwf.org.uk/ (accessed on 02 March 2009)
[81] The IWF remit also covers obscene content and racial hatred
[82] The exact design of CleanFeed is not in the public domain
[83] An advocacy and lobbying organisation http://www.chis.org.uk/ (accessed on 01 June 2010)

Zoe Hilton [10] *NSPCC* policy adviser claims using *CleanFeed* (or similar measures) is important, despite their limitations, to prevent accidental exposure to child abuse and paedophilia sites. A common argument given in support of *CleanFeed* and other censorship is to prevent accidental exposure to paedophilic material, as this can trigger a descent into child abuse [12]. This point has been raised and challenged in parliamentary discussions [13] and has even been disputed by seasoned industry experts [14]. Richard Clayton [11] disputes the likelihood of accidental exposure to such material, claiming that most of material is on paid-for sites or privately held in closed networks. Publicising or indexing child abuse websites would be necessary for any realistic[84] accidental exposure and would quickly lead to content take downs and arrests for content hosts and providers.

Although *CleanFeed* is probably the best known, other technologies (from other companies) exists to implement the *IWF* blacklist.

Australian Government plans for a similar network level censorship programme have met with opposition from academia, civil liberties organisations and children's charities [89]. Opposition to the blacklist is based on the following:

- Efficacy of the filter – Simple to bypass, only covers http traffic
- Civil liberties – Lack of transparency is open to abuse and mistakes
- Funding[85] – Money would be better spent if diverted to child protection authorities, prevention and education programmes.

Around half of the Australian blacklist[86] has nothing to do with child abuse imagery [88].

5.8.5 Community Reporting and 'Safety Mode' – Learning from YouTube

YouTube[87] is a video sharing website created in 2005 which allows users to upload video – these videos are (automatically) encoded to a *Flash*[88] format for streaming to a web browser. Every minute, approximately 10 hours of video is uploaded to

[84] Discounting randomly typing URLs or IP addresses into a browser.
[85] The Australian blacklist is to be tax payer funded, IWF is a charity.
[86] Unlike the IWF, the Australian blacklist has been leaked into the public domain by wikileaks.
[87] http://www.youtube.com/ (accessed 11 Feb 2010).
[88] http://www.adobe.com/products/flashplayer/ (accessed 11 Feb 2010).

YouTube [38], moderating this much content is un-scalable. *YouTube* relies on users to flag inappropriate material. Flagged videos are reviewed and removed if they violate *YouTube's* 'Terms of Use' – flagged videos are not automatically removed [39].

To augment this process *YouTube* launched "Safety Mode" to shield children and other vulnerable individuals from inappropriate content that does not violate its terms e.g. graphic war coverage. Additionally it collapses (hides) comments by default and censors profane words with asterisks [39].

This pragmatic approach allows inappropriate content to be flagged and removed without the resource overhead typically associated with moderation; it also allows censorship for children without impacting the entire community.

5.9 Security Theatre

Security theatre is a term coined by Bruce Schneier to describe security measures that superficially appear to improve security but in reality provide little or no improvement in security or reduction in risk.

In his blog [24] Schneier describes the use of Radio Frequency Identification (RFID) tags at hospital maternity wards. The infant abduction rate is tiny, an order of magnitude below infant mortality yet infants are equipped with RFID anklets that trigger an alarm if they pass through scanners on the maternity ward doors. If an infant is removed from the ward an alarm goes off.

The risk of infant abduction is very small and the increase in security from RFID tagging of infants is also very small. Schneier notes that RFID technology provides little real security but praises the approach for providing parents with a sense of reassurance and peace-of-mind.

Many online child protection measures are questionable in their efficacy [7] [9] [25] and might be classified as security theatre. However, can they be considered worthwhile if they reassure and bring the parents risk perception in line with genuine levels of risk children face on the Internet. Furedi [3] has suggested that paranoid

parenting and a culture of fear are retarding child development and stifling society. One of the benefits of security theatre in this domain may be to reduce these fears in parents.

Beginning in April 2009 paedophiles and other sex offenders on probation in the West and East Midlands National Offender Management Service (NOMS) area will face mandatory polygraphs[89]. During this three year trial period 600-1000 offenders will face regular examination. Those refusing to participate will have their licence revoked and be returned to prison. It is expected that successful trials will see polygraph tests extended to other regions and other crimes [26]. Are polygraphs an effective tool for monitoring sex offenders? Within the scientific community polygraphs are regarded with scepticism – there is no common baseline for deception (lying) and the results prove to be highly variable (unreliable), averaging only slightly better than chance [27] [28].

This suggests subjecting offenders to polygraph is not an efficient use of resources in the child protection ecosystem. Looking beyond the empirical data, what other effects does the polygraph have? The perceived accuracy of the polygraph may encourage voluntary disclosures on the part of the test subject (the sex offender), that may have gone otherwise undetected in the examination. The knowledge that the subject will undergo a polygraph where any violation (recidivism) would be detected may act as a deterrent (against offending). Here perception of the efficacy of the polygraph maybe also important. These points too are open to question. Polygraph use by US intelligence agencies (which is common), shows no evidence of stemming disclosures of classified information (leaks) [29].

It is reasonable to assume sex offenders are subject to polygraphs to monitor and defend against recidivism. How prevalent is recidivism among sex offenders? According to the US Department of Justice, Bureau of Justice Statistics, the recidivism[90] rate for sex offenders in general is 5.3% within 3 years and 3.3% for child molesters [30]. It also found 40% of the alleged sex crimes occurred within 12

[89] Sometimes referred to colloquially as ‘lie detectors’.
[90] Defined as rearrest, reconviction and reimprisonment during the 3 year follow up period.

months of offender release. *Unofficial figures*[91] *suggest the recidivism rate may be up to 4 times higher, higher still if behaviours and thought crimes are included* [94].

Given the low rate of recidivism and questionable efficacy of polygraphing, is polygraphing sex offenders on licence a cost effective measure? There is however another facet to the question, if parents and society at large believe polygraphing sex offenders makes children safer, provides them peace of mind and helps bring the overall perception of risk into line with actual risk, should this be included in the argument in favour of polygraphs? As with RFID tags in maternity wards, the qualitative feelings of peace of mind and re-alignment of risk are difficult to quantify – how can a monetary value be applied to them?

It is beyond the scope of this thesis, but the author believes a cost / benefit analysis of polygraphing sex offenders, compared with other offender management processes should be explored. Forensic psychological interviews to determine whether an offender should be released on licence, closer police / probation service monitoring on release and other treatments could all be explored.

5.10 Realigning risk

Parents are perhaps most fearful of their children being duped into meeting someone they 'know' from the Internet and being forcibly molested. As has already been discussed this is highly unlikely, the majority of minors who do meet adults and have sex, know they are meeting an older man for the purposes of having sex. This fear has left parents with a disproportionate perception of risk; this perception has likely been fed by egregious media reporting and anecdotal accounts. When examined quantitatively how does the risk from grooming compare to other risks to children?

- The British Crime survey[92] records 314 incidences of grooming in 2008.
- The Department for Transport[93] reports 2087 killed or seriously injured children on UK roads (124 died) in 2008.

[91] Some of the data sources are of low quality – the Police lost track of 19% of offenders.
[92] http://uk.sitestat.com/homeoffice/rds/s?rds.hosb1109chap2newxls&ns_type=clickout&ns_url=[http://www.homeoffice.gov.uk/rds/pdfs09/hosb1109chap2new.xls] (accessed 03 Dec 09).
[93] http://www.dft.gov.uk/pgr/statistics/datatablespublications/accidents/casualtiesmr/rcgbmainresults2008 (accessed 04 Dec 09).

- Cancer Research UK[94] states there are around 1,500 new cases of childhood cancer diagnosed each year in the UK.

From these figures it can be deduced that children are more than six times more likely to be killed or seriously injured on the road and nearly five times more likely to be diagnosed with cancer. According to the Office of National Statistics[95] there are approximately 12.3 million children, this gives an unqualified risk of around 5 reported incidents of grooming per 200,000 children.

Note on the above statistics: As has already been discussed, teenagers are at much greater risk than younger children and girls are at greater risk than boys. Of the 314 reported incidences of grooming there is no breakdown of this figure into online/offline. Also it should be noted that there is the possibility of grooming incidences going undetected/unreported, whereas it is unlikely that a seriously injured child or child diagnosed with cancer would go reported.

Would parents be more relaxed, (have greater peace of mind) if their perception of risk[96] more accurately resembled actual risk? How can parents' perception of risk be realigned to more closely resemble actual risk?

The media needs to maximise ratings (for profitability) and often does this through sensational and alarmist reporting, opportunistic politicians and special interest advocacy groups. Excessive parental fear or risk probably stems from these reports [109] [110]. This excessive fear is at best unhelpful and probably counterproductive [3]. In a free society, barring exceptional circumstances, freedom of press however unhelpful cannot be curtailed. However, the government and official sources (such as CEOP) should attempt to advance a responsible media agenda to help parents understand the real risk and rebalance their fears.

[94] http://info.cancerresearchuk.org/cancerstats/childhoodcancer/incidence/index.htm (accessed 10 Dec 09).
[95] http://www.statistics.gov.uk/cci/nugget.asp?ID=6.
[96] Assumes parental risk perception has been distorted.

5.11 Trade offs and Best use of Resources

No system can ever be completely safe or secure, human nature means there will always be crime and there will always be dangers to children. It may seem unpleasant to some, but society must accept that some children will be solicited, molested, bullied, exposed to unwanted content and come to harm.

Security is a process of compromises [33]; it may by possible to improve Internet child safety by universal monitoring of Internet connections by ISPs or some other official authority. Ignoring the technical feasibility, society would probably oppose this on civil liberties grounds. Parents could forbid their children from using the Internet, even assuming their children obey, this would be cutting them off from an enormous educational and social resource, likely having a negative impact on child development [3].

With finite resources researchers must evaluate how effective child protection measures are. This is a relatively simple task when determining the efficacy of technical measures such as content filters; a quantitative metric can describe the performance of such measures. This only describes the performance of the technology, nothing about the benefit to the child - *how damaging would the exposure have been? Is failing to block a site about drug use worse than failing to block a pornographic website?* Social strategies such as education are harder to evaluate.

Psychology researchers need to define qualitative metrics to describe the benefits of social engagement strategies such as education and training. Until such metrics exist a cost-benefit analysis cannot be performed and a comparison between strategies and technologies is not possible. Such work will undoubtedly be subject to a certain amount of bias and require a degree of expert speculation, but would probably be of value to decision makers and budget holders.

From 2009 the Independent Safeguard Authority[97] (ISA) was established to register and vet individuals working with vulnerable groups. This could see 1-in-4 adults

[97] The ISA decides whether or not an individual is fit to work with vulnerable groups, a CRB check discloses offences. A CRB check is part of the ISA vetting process, but can be performed independently for other reasons.

being vetted [103] as part of their work or social life. In 2008/2009 2,551 disputes were upheld against the Criminal Records Bureau[98] (CRB) [95], it does not seem unreasonable that more false negatives would likely go unnoticed (people would investigate and dispute false positives but not false negatives). Enhanced CRB checks include soft intelligence (accusations, suspicions, investigations that did not result in conviction) as well as criminal convictions. Enhanced CRB checks are required for jobs where people have regular access to children. This is likely to exclude a significant number of people based on false positives and is obviously open to malicious abuse [96]. Seven UK organisations representing head teachers have described the process as "disproportionate to risk" [97]. Sir Roger Singleton, chair of the ISA has warned against over reliance on the database [98] "Frankly all registration means is that a) you've paid the fee and b) there is no known reason why you should be on one of the barred lists."

Overlooking these shortcomings and limitations the vetting process is designed to improve child safety, even if it is draconian and disproportionate. It is not possible to conduct a controlled study to see if children are safer and the low rate of crimes against children would likely mean any reduction was small, perhaps even statistically insignificant. Loss of potential employment and potential scandal suffered by those subject to failed vetting due to false positive is not the only negative side effect of the vetting process. Some children's authors have refused to be vetted and thus refused to appear in schools [99]. The voluntary sector is suffering as people are put off volunteering by the process of vetting or the current child protection climate [100].

Assuming the vetting process does indeed make children safer, does the additional safety make up for the loss of volunteer run programmes? Such programmes may involve sports, citizenship (e.g. scouts), or general youth clubs. The loss of these programmes *may* have unwanted side effects, including potentially increased delinquency, anti social behaviour and substance abuse. Which is of greater benefit to society, (potentially) improved safety or volunteer organisations and youth groups? Furedi and Bristow [100] are in no doubt that the current process and 'social atmosphere' that has been created is to the detriment of society as a whole.

[98] A CRB check is part of the ISA vetting process.

This is only one example of the trade offs that must be made, parents must also consider the benefit of stranger contact (rewarding relationship) versus the risk of harm. As described, high risk activities such as sex chat with strangers has benefits that must be considered. Even the act of using filters, monitoring and rules need to be contrasted with the requirement to expose youth to a certain amount of risk in accordance with individual development.

5.12 Difficulties Posed to Researchers

The evaluation of technical protection measures is a relatively straight forward task that can be performed using standard scientific methods. These measures cannot be tested in isolation. From a systems perspective, people are the most significant component in this problem domain, it is essential to examine the human factors and stakeholders.

Stakeholder analysis is the backbone of requirements capture and understanding the culture surrounding a system. A typical stakeholder analysis would include interviews, questionnaires and observations. The most prominent stakeholders in online child protection are children, parents, service providers, law enforcement and child abusers. Stakeholder analysis of two of these groups' children and child abusers is fraught with difficulty.

The nature of the subject would necessitate questions involving violence, sex and personal history. Among post pubescent minors, even if their parents are happy for them to engage in the process, a small but significant number are likely to be dishonest [104]. Questioning prepubescent children on such subjects is, at best, unethical and possibly illegal as it risks exposing them to information beyond their understanding and maturity level.

Child abusers and paedophiles not already in the criminal justice system are unlikely to participate – it would likely entail the admission of illegal activities or, at least 'socially unacceptable' behaviours and urges. Young conducted forensic psychological interviews on 22 sex offenders [105]. While her findings are not in dispute, the sample size is small and no questions appear to have been asked about the

motivations of participants. Those in the criminal justice system may lie during the interview process, perhaps being in denial, attempting to minimise or mitigate their guilt, or attempting to blame-shift [14]. They may even have been feigning cooperation in an attempt to curry favour with prison authorities. Such issues must be considered when weighing the value of the results.

The tabloid press has no interest in rational debate [109] surrounding child abusers and paedophiles; it regards them as monsters that require punishment [2]. A systems approach requires an unbiased consideration of stakeholders and their actions - emotive knee-jerk rhetoric and 'comfortable' stereotypes ignore that children are often active participants in their own abuse, indeed society does not want to believe such actions are true [14]. Researchers who question this ingrained mindset risk being labelled unethical or 'playing into the hands of child abusers'. Researchers have reported organisations unwilling to cooperate with them because of the potential for negative PR and reduced funding – one university refused to publish a PhD thesis on these grounds. Much of this reticence is believed to stem from a fear of the treatment individuals and organisations will receive in the tabloid press [106].

Despite this climate some organisations and individuals are conducting probing research including the *Crimes Against Children Research Center*[99]and danah boyd[100] [sic]. However even this work (through the ISTTF [4]) has drawn public criticism from the Attorneys General of Pennsylvania and North Carolina [5] as discussed earlier. The AG have been thoroughly rebuked by academics [20] [21] for political grandstanding and fear mongering.

Such is the breadth and sensitivity of the subject a meta-analysis has been necessary to gain an holistic perspective of the problem space, particularly with the fear of negative PR. Government and law enforcement statistics (sometimes at odds with their public message), journal articles on adolescent development, psychology and medicine, parliamentary questions, testimony and enquiries, expert opinion and observation, even social commentary all provide valuable source information when collating a systems view. Researchers should evaluate all sources with a critical

[99] http://www.unh.edu/ccrc/index.html (accessed on 27th January 2010).
[100] http://www.danah.org/ (accessed on 27th January 2010).

demeanour, particularly with government or other potentially biased sources which might be (perhaps for political reasons) improperly reported, analysed or subject to withheld methodologies [107] [108].

5.13 Discussion

Recent years have seen a wholesale change in attitudes toward online child protection, as discussed earlier, possession of well known Internet images depicting members of the *Simpsons*[101] cartoon family in lewd poses saw two separate convictions in Australia. One in 2008 for 'possessing child pornography', the other in 2010 for 'possessing child exploitation material' [52] [53]. Possession of these images would also be a criminal offence in the UK (under s49 of the *Coroners and Justice Bill*) [47]. In 2000 staff at *Royal Sun Alliance,* were disciplined (10 were dismissed) for distributing the obscene *Simpsons* cartoon by email. Commentators at the time called for restraint and common sense *"Porn, especially child porn, is one matter, but smutty cartoons? Come on."* [54] and *"Are they really that offensive? Not really…"* [55]. Had the emails been sent today, the perpetrators would likely have attracted fines, prison sentences and been made to sign the sex offenders register. Would this have made children any safer? Do laws such as this improve child safety? The author feels not, ignoring the absence of a proven causative link, such material could be evaluated (and legal proceeding pursued) under the Obscene Publications Act against the 'tendency to deprave and corrupt' test of obscenity [60].

Politicians and public officials have a duty to serve the public, using child protection as a platform for political gain is unhelpful [21] [41] [84] [86], it serves to confuse the public and distract from valuable work being undertaken in the problem space.

5.14 Conclusions and recommendation

The 2002 Soham murders pushed child safety front-and-centre in the public mind. As Internet access has surged towards ubiquity in the home and now on personal devices, parents have become understandably concerned about new risks. There are undoubtedly individuals who abuse children and use the Internet to help facilitate this

[101] It is not known if the lewd *Simpsons* cartoons were an identical set, it thought they are part of the same superset.

abuse; such individuals represent a small minority of child abusers. Those who do meet and have sex with minors are rarely deceitful or forceful; their victims are typically physically mature, sexually receptive teenagers. They groom their victims with attention and affection, behaviour which, if not for the age of consent, would be considered dating. Victims frequently refuse to help authorities prosecute offenders, citing feelings of love and affection. Parents are understandably worried when the media focuses on the most alarming cases and are likely overwhelmed by the array of safety advice and products and services to protect children. Their concerns and confusion are not helped when some official advice is known to be unrealistic and products that are designed around technical aims ignore the way children behave.

While one-in-seven children are solicited this is not the same as grooming, it is often a single off hand comment in a chat room. The majority of children deftly handle such encounters and report no negative side effects. Cyberbullying and harassment is more than twice as common as solicitation and more children report distress resulting from bullying than solicitation. Cyberbullying may even have greater long term ramifications than real world bullying.

Most people are keen to prevent children's exposure to inappropriate content, particularly pornography because of its perceived harmful effects. Withholding pornography from children is not an unreasonable moral position, but evidence does not support claims of harmful effects. Computer generated and other artificial child pornography has been outlawed in the UK without evidence that it causes harm; it may in fact be increasing the risk to children.

Technology has a part to play in keeping children safe on the Internet but it is a minor role and parents should not be reliant on it. It's most effective at protecting younger children from accidental exposure to inappropriate content. At least some teenagers are actively seeking such content and may be able to circumvent technological measures or simply use a public terminal or friend's computer. Ignoring teenagers as stakeholders, failing to address their culture with a systems mindset, will continue to see products and services at best under perform and at worse do more harm than good [101] and risks a false sense of security.

Social measures, education and close well formed relationships between parents and children and oversight are major factors in keeping children safe. Education applies to children, parents and others with an interest in child safety. Children should receive age appropriate information about risks and how to protect themselves, *Thinkuknow* is a very good source but needs to be supplemented in schools with information that reflects how children behave and should include role playing what to do when faced with solicitation, bullying and online exposure. Parents and others involved in child welfare need to be aware of the various technologies so they can communicate and relate to children. This needn't be complex; knowing that IM is a chat environment and what a social network profile is and how important these things are to children are invaluable in understanding a child's world. This understanding and an accurate perception of the risks rather than fear stoked by media scare stories help build a strong relationship. Testing boundaries and engaging in at least some forbidden behaviour is to be expected with teenagers. A strong relationship will mean those teenagers will turn to parents, teachers or social workers if they become involved in a serious situation or are unable to cope.

Without universally recognised definitions and metrics side-by-side analysis of research is difficult and the quality of results questionable. Until such standards are agreed, the literature and the broader understanding of the problem will vary widely. Without this, the requirements for programmes, products and services risk being inaccurate and incomplete. Without adequate requirements capture, definition and validation any solution to a complex problem will be left wanting.

5.15 Business Impact

The systems analysis caused the author and many he has spoken with to challenge their preconceptions about the online child protection problem space and how best to address its challenges. BT currently provides Cleanfeed, an ISP level content blocking system, however the 'real world' systems value of the IWF blacklist has been questioned - the chance of accidentally finding genuine child pornography is incredible low (there are between 500-1200 web pages that are not indexed among over 200 million websites [112] and in excess of one trillion indexed pages [113]). However such security theatre can be merited if it provides reassurance to the public at large. BT also provides its customers with content filtering software from McAfee

free of charge, the author believes this too be a valid action if for no other reason than it keeps BT broadband package competitive in the marketplace.

This work largely indicates that BT should not engage other technical methods to protect children on the internet e.g. network level content analysis or predator detection. Though it should of course maintain a research and analysis interest in technical protection methods.

The best way to protect children online is through education of both parents and children. BT has the skills, experience resources and trusted brand to engage with parents, children, teachers and other stakeholders in outreach and education programmes. Such engagements could form a part of BT's Corporate Social Responsible (CSR) agenda. The value of such programmes cannot be readily discerned from a balance sheet but would almost certainly have a positive impact on BT's brand and reputation, generating goodwill with existing customers and perhaps winning new business.

5.16 Reflective Commentary

The Online Child Protection work evolved from work to establish whether is it possible to detect sexual predation based on text based chat. As a side issue the author begin to examine the wider issues of the problem space in order to satisfy the systems requirements of the doctorate. The author was initially of the opinion that many of the challenges could be addressed with technology, the systems approach lead to a complete reversal of this opinion

An analysis of the stakeholders, the interfaces, an understanding of dynamics of abusive relationships, and human factors in particular developmental normal teenage behaviours, lead to the realisation that technology is only a bit player in the protection of children. It is important, but the return of investment is inferior to proper education, training, strong relationships and good support structures.

Content filters and some other products designed to protect children are valuable in shielding younger children from some of the potentially harmful recesses of the Internet. However such measures can be readily circumvented even without any

technical skill e.g. using a public terminal or friend's computer. Teenagers are headstrong individuals beginning to assert their identity and seeking privacy - some boundary testing and defiance is to be expected. Unrealistic expectations and draconian rules are unlikely to be followed, rules need to be based on consent and understand ('user buy-in').

In addition to the author's own perceptions the subject of child protection is emotive and highly political. Many of the approaches to child protection amount to security theatre that do not perform as intended 'outside the lab', some politicians and the media are using the subject publicity and not to advance the cause, all of which have lead to a distortion of the problem space. The systems thinking approach dramatically changed the author's perception of the problem and he considers it should be essential practice to anyone working in the field.

5.17 Systems Analysis Report
The following report, has been circulated within BT. It is an abridged version of the research and findings from this chapter.

A Systems View of Online Child Protection

Abstract
Online Child Protection is never far from the public conscience in the UK as parents and society at large seek to keep children from harm on the Internet. There is much fear, uncertainty and doubt as parents are exposed to mawkish anecdotes in the media, and receive erroneous advice from industry and official sources. This paper takes a systems view of the problem space - it examines the size and nature of the problem, what measures exist to protect children, whether or not they are effective, and what factors affect the risk to children online. It finds the traditional protection message and advice is based on unsubstantiated assumptions, is widely ignored, and may actually be unhelpful. It also describes the challenges to researchers, and discusses how close parent/child relationships, and education both within the home and school environment are most effective ways to protect children online.

Keywords
Online Child Protection, Internet Child Protection

Authors:
Chi Durbin
Mark Pawlewski
David Parish

ISSUE Issue 1.0
DATE 29th April 2010

Document Information

Title	A Systems View of Online Child Protection
Owner	Chi Durbin
Author(s)	Chi Durbin, Mark Pawlewski, David Parish
Location of electronic copy	
Location of paper copy	None
Change Authority	

Change Control

Distribution

Issue	Author	Date	Details of Change

Confidentiality Statement

Copyright

1 Introduction....128
2 Systems Thinking....128
3 Paedophiles, Child Molesters and Online Grooming....129
4 Where are the dangers? What are the risk factors?....130
5 Current Protection Methods....132
6 Difficulties Posed to Researchers....133
7 Discussion....135
8 Conclusion....136
9 References....138

INTRODUCTION

In August 2002 two, ten year old girls were murdered in Soham, Cambridgeshire (UK). Subsequent media coverage and government enquires have elevated the subject of child protection to prominence in the public arena. New rules for those interacting with children could see 1-in-4 adults being vetted [Ozimek, J., 2008, 14 July] as part of their work or social life.

During the same period worries about child safety on the Internet have risen. This is driven in part by alarmist television shows such as *To Catch a Predator*[102] but also by education and awareness campaigns pushed out by government and industry. This is often bundled with advice regarding Internet security and identity theft.

Internet usage in the UK is currently (May 2009) running at 76% [103] (of the population), this presents a new and relatively safe channel for abusers wishing to approach children. Furthermore, the explosion of Social Networking Sites (SNS) and blogs provides a potential source for 'child-mining' – with children having uploaded photographs and personal information. Here the public perception is of child molesters using information gleaned from such sources to manufacture online personae to mask their age and sexual intentions, grooming naïve young victims and enticing them to offline meetings [Wolak, J., Finkelhor, D., Mitchell, K., & Ybarra, M., 2008].

The reality of sexual predation on the Internet is somewhat removed from this perception. Children who meet Internet groomers are overwhelmingly physically mature teenagers, who know they are meeting older men and intend to have sex with them. Many do so more than once [Mitchell, K., Finkelhor, D., & Wolak, J., 2001], [Smith, A., 2007]. However, Internet initiated grooming of sexually immature, prepubescent children is almost non-existent [Internet Safety Technical Taskforce, 2008], [Wolak, J., Finkelhor, D., Mitchell, K., & Ybarra, M., 2008].

Looking beyond what shrill elements of the media and industry are telling the public, this paper examines the problem space from a systems perspective. Describing the application of systems thinking – stakeholder analysis, threat vectors, risk factors, behaviours, and existing protection techniques. It describes the difficulties researchers face in such a sensitive arena, and discusses what the authors' believe to be the most effective way to protect children, and highlights areas where additional work is needed.

SYSTEMS THINKING

Systems Engineering seeks to manage the complexities inherent in modern engineering projects. It encompasses processes, technology, and management activities to define systems, capture and verify requirements, validate designs, manage conflicts and trade-offs, develop architectures and co-ordinate disparate specialist product teams. Its mission is to deliver a capability that satisfies the customer requirements, using the least resources, at the best cost.

The traditional techniques, methods, models and metrics of systems engineering may not, at-a-glance seem applicable to the online child protection ecosystem. However there is much value in the systems ideology of 'stepping-back' and taking a holistic view of the problem space to understand factors such as:

- Stakeholder requirements and motivation (abusers, children, parents, law enforcement, prisons and reform services , service providers, society at large)

[102] http://www.itv.com/News/tonight/episodes/Tocatchapredator/default.html & http://www.msnbc.msn.com/id/10912603 (accessed 15 Oct. 09).
[103] http://internetworldstats.com/europa.htm (accessed 12 Jan 10).

- Conflicting requirements and trade-offs (privacy vs safety)
- Threats facing children on the Internet (grooming, bullying etc)
- Factors and behaviours that increase risk
- Perceived risk vs actual risk
- Existing protection measures (efficacy and limitations)
- Legacy requirements (limitations of the Internet, and existing services)

This list is only a limited sample of the many considerations, and is presented to give the reader a brief idea of how systems thinking can be applied.

PAEDOPHILES, CHILD MOLESTERS AND ONLINE GROOMING

The term 'paedophile' is emotive and divisive. It is often bandied around in the media as a catch-all for anyone with a sexual interest in minors. It is an ambiguously defined and poorly understood label. The *American Psychiatric Association* in *The Diagnostic and Statistical Manual of Mental Disorders* (DSM)[104], specify the following diagnostic criteria for paedophilia [American Psychiatric Association, 2003].

D. Over a period of at least 6 months, recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving sexual activity with a prepubescent child or children (generally age 13 years or younger);
E. The person has acted on these sexual urges, or the sexual urges or fantasies cause marked distress or interpersonal difficulty;
F. The person is at least age 16 years and at least 5 years older than the child or children in Criterion A.

The general consensus is that paedophilia describes adults with a sexual interest in prepubescent children, and where the adult is at least five years older than the child. This excludes so called *Romeo and Juliet* romances between peers where one individual is under the age of consent. In the UK the *Sexual Offences Act 2003*[105], sexual acts between minors are outlawed but government advice to Police and the Crown Prosecution Service is not to actively pursue these cases [BBC, 2004]. The terms hebephilia and ehebephilia describe adults with a sexual interest in pubescent and post-pubescent minors.

Lanning [Lanning, K., 2001] discourages the use of the term 'paedophile', cautioning that it is a (psychiatric) diagnostic term that describes desires and interests not intentions or actions. The term 'child molester' more accurately describes the sexual abuse of children.

Lanning argues child molesters can not be easily placed in discreet categories, instead falling along a continuum, from situation sex offender to preferential sex offender. Briefly, situational sex offenders are likely to be less intelligent, opportunistic, impulsive, and follow modus operandi (MO) behavioural patterns. Preferential sex offenders (at the opposite end of the continuum) typically exhibit more intelligence, focused criminality, compulsion, and ritualised patterns of behaviour.

Groomers are individuals who befriend children, with the intention of gradually engineering the 'friendship' towards a sexual relationship. This would begin with fairly benign behaviour such as playing with the child or buying them an ice cream. Gradually the groomer would introduce the child to sex, typically with exposure to pornography and child pornography to normalise the subject within the child's mind. Simultaneously the groomer would ingratiate

[104] Commentators have questioned the listing of paedophilia and other paraphilia (psychosexual disorders) in the DSM, suggesting they been listed without any scientific or rational basis (Tromovitch, P., 2009).
[105] http://www.opsi.gov.uk/Acts/acts2003/ukpga_20030042_en_1 (accessed on 15 December 2009).

themselves with the child, showering them with attention, buying them gifts, telling them they were special. Sexual activity may begin with kissing and touching. The child will be told this is normal behaviour and maybe exposed to other (already) groomed children. The child may be told to keep the activities a "special secret" or coerced into silence through threats, or told the groomer would get into trouble if people found out.

Online groomers are often described as using similar tactics but hiding behind a fake persona – typical that of someone in the child's peer group, often slightly older and more 'worldly'. By the time they meet the child, the relationship with the child will be strong enough to withstand any sense of shock or deception the child feels. The reality of online grooming does not reflect this perception.

Evidence indicates that far from concealing their identities and intentions groomers are open and candid, revealing both their ages and sexual interests in minors during Internet encounters [Wolak, J., Finkelhor, D., Mitchell, K., & Ybarra, M., 2008]. In the majority of these encounters the victims were active, even enthusiastic participants in their abuse. They may even have initiated it. Society struggles to understand that these victims are post pubescent adolescents - physically mature and sexually curious; it refuses to accept they are anything but innocent participants forced against their will [Lanning, K., 2001]. This misconception is reinforced in the media and sometimes official information. This may be hindering public advice and leaving parents ill informed and poorly equipped to protect their children. In fact online solicitation of pre-pubescent minors is extremely rare [Internet Safety Technical Taskforce, 2008]. Younger children are more closely monitored by parents, less available online than teenagers, and on the whole not interested in sex and romance. This makes them a much smaller target [Wolak, J., Finkelhor, D., Mitchell, K., & Ybarra, M., 2008].

WHERE ARE THE DANGERS? WHAT ARE THE RISK FACTORS?

What are the real dangers facing children on the Internet? Do certain activities, applications, and platforms pose a greater threat to children than others? What behaviours on the part of the children increase the risk they face? What other factors are involved?

A common activity of children on the Internet is the use of chat rooms [Ybarra, M., & Mitchell, K., 2004], this is also the most common (76%) venue for first encounters of Internet initiated sex crimes [Mitchell, K., Finkelhor, D., & Wolak, J., 2001]. Typically the chat rooms were keyed towards teens, geographic locations, dating and romance, and worryingly, in a few cases, sexual encounters between adults and minors. Of the victims of Internet initiated sex crimes, 75% were aged 13-15, 1% were aged 12, none was under 12. 75% of victims were female, 96% of offenders were male. While it is clear that discussing sex and relationship in chat rooms, particularly with strangers' carries risk above that of normal Internet use, it also has developmental benefits that may not be immediately apparent.

Subrahmanyam et al [Subrahmanyam, K., Greenfield, P. M., & Tynes, B., 2004] on sexuality and identity in teen chat rooms found that peer group communication is the number one source of sexual information for adolescents (though little is known of what they are telling one another). Adolescents talk about sex online as a way to understand and control their feelings, maintaining open communications can serve as a coping mechanism for sexual expression, particularly when providing an anonymous forum for discussing embarrassing issues. It maybe preferable to have children engage in online relationships and cybersex[106] as this entails less risky behaviour, particularly unprotected sex. Cybersex and online relationships may allow girls (the greater risk group) to exercise more control as they are recognised to possess greater communication skills.

[106] Cybersex is a form of computer mediated communication where users send each other sexual explicit messages – it may include fantasy, role playing, and real life masturbation.

Social networking has seen a meteoric rise in recent years, in January 2009 the two top Social Networking Sites (SNS) - *Facebook* and *MySpace* recorded over 2 billion visits [McCarthy, C., 2009]. While the format, layout and target demographic of the various SNS platforms differ, a user typically posts information and photos of themselves and links their profile to friends. Some SNS profiles take advantage of privacy controls to limit the exposure of information to the larger Internet, others do not.

Does the use of SNS and having a profile increase the risk to minors online? Research at Pew Internet [Smith, A., 2007] found that 49% of teenagers use social networks to make new friends and 32% were contacted by a stranger[107]. 7% have been left scared or upset at one time, by contact with a stranger, this was far more likely among girls (possibly because of egregious media information). No evidence was found that linked the posting of personal information (real name, address, email, school) to stranger contact. This repudiates the notion of groomers harvesting information to spot victims. Those who posted photos online were more likely to be contacted by a stranger. It should be clear that stranger contact is a not necessarily negative, mostly it will be positive or at the least benign [Livingstone, S., & Helsper, E., 2007] – the majority of strangers (adults and minors) have no ill intentions. Further, it is important to realise youths widely ignore warnings not to communicate with strangers, seeing the advice as unrealistic [Mitchell, K., Finkelhor, D., & Wolak, J., 2001].

The majority of child abuse takes place within the family or by individuals well known to the family (friends, neighbours, teachers etc). Offline abuse, along with neglect and disadvantage increases the risk or likelihood of the victim becoming involved in Internet initiated abuse. All types of real world abuse (physical, psychological, sexual), poor parenting, lack of supervision, alcohol and substance abuse, poor relationships between parent and child, mental health problems, bullying and peer-group isolation are all factors that increase a child's risk online [Internet Safety Technical Taskforce, 2008]. Being female, frequenting chat rooms and discussing sex online especially with strangers are also associated with increased risk [Wolak, J., Finkelhor, D., Mitchell, K., & Ybarra, M., 2008].

This paper deals mostly with the risk to children from grooming, largely because this is the greatest fear of parents, however it is not the only threat children face on the Internet. The two other main concerns are exposure to unwanted material and bullying.

Although parents may wish to shield their children from violence, hatred, drugs etc the main concern is pornography. There is little evidence to suggest children suffer any harm from exposure to pornography and virtually no research on the matter [74]. Although a minority of children were upset by unwanted exposure to pornography most considered it an irritant, like spam and felt in it was the 'price of doing business on the Internet'. It is not known how 'upset' this minority were from their exposure. Younger children often lack the maturity to understand pornography and are simply annoyed by unwanted exposure, they may however be distressed by extreme pornography – explicit scenes of bondage and sadomasochism where pain and physical trauma are involved.

Online bullying has no consistent definition and covers an array of behaviours including impersonation, outing (revealing secrets), denigration, exclusion, and harassment. Online bullying shares many attributes with real world bullying but includes the possibility of the bully being anonymous. It is estimated that around 30% of children are involved in online bullying either as an aggressor, a victim, or both, with 19% regularly involved in some form of bullying [Wolak, J., Mitchell, K., & Finkelhor, D., 2003]. It was found that the use of Internet rules, and blocking software had no significant effect on bullying. 30% of victims felt upset, 24% were afraid, and of those that were neither upset nor afraid 34% still experienced

[107] Defined as someone with no connection to you or any of your friends.

some effects such as aversion to the Internet, feeling jumpy or irritability [Internet Safety Technical Taskforce, 2008].

CURRENT PROTECTION METHODS

What methods exist to provide a measure of protection for children on the Internet? To enable the reader to better understand the issues, some of the key methods are discussed here.

Numerous software packages offer tools for parents to install on their home computers. These typically allow parents to block certain applications, filter websites and review online activities. Blocking is often unwieldy as some legitimate use is often permitted. The filtering of websites consists of a list of blacklisted (or whitelisted) URLs or dynamic analysis of content. With the continuously evolving nature of the Internet and its massive size, manual evaluation of sites is un-scalable. Filtering of any type is subjective – at some stage an individual, committee or other authorising body will have defined what is acceptable, what is not, and create categories (adult content, violence, hate groups, etc). The subjective nature of filters, their general efficacy, and the claims of providers have all been brought into question [National Research Council, 2002].

Reviewing technologies vary between applications, but may include website histories, email and IM logs, keyloggers (recording every keystroke) and even replay of video calls. In a family with more than one child or where a child spends more than brief periods online, this is prohibitive in terms of the parents' time. Older children may decry these as being excessive and an invasion of privacy.

In 1996 the *Internet Service Provider Association* (ISPA) created the *Internet Watch Foundation* (IWF)[108] to catalogue child abuse[109] websites, this is compiled into a blacklist which is provided to ISPs to block access. BT CleanFeed[110] is a popular ISP level filter - traffic destined for a suspicious IP address is diverted to HTTP proxy where the destination is compared to the IWF blacklist [Clayton, 2005]. The raison d'être behind the IWF is to prevent accidental exposure to child abuse (the viewing of which would motivate real life abuse), however this has been disputed [Cellan-Jones, R., 2009] on the basis that most child abuse websites are pay sites or privately held and indexing this information would lead to exposure and arrest. Also questioned is the commonly touted notion that exposure to such images triggers abuse [Lanning, K., 2001]. The IWF blacklist is updated twice daily and usually contains 800-1200 live sites with 50 URLs added daily [Ozimek, J., 2008, 9 October], as of September 2009 the blacklist contains around 500 URLs (a near all time low) [Ozimek, J., 2009]. Around 95% of UK ISPs employ CleanFeed or a similar product to implement the blacklist, these products can be easily circumvented by using international proxies.

Many images on the Internet are altered (sometimes referred to as 'Photoshopped'), indeed photo realistic images can be computer generated without the need for a human subject. With no victim, where no child is harmed, the only rational argument against computer generated child pornography is that it leads to a normalisation of inappropriate feelings, which in turn leads to an increased risk of (real life) child abuse. This is the generally accepted view [Ozimek, J., 2009, 17 March] [TheyWorkForYou.com. 2009] [Carr, J. 2003]. Though it enjoys first amendment protection in the US, all artificial child pornography is prohibited in the UK under section 49 of the *Coroners and Justice Bill*[111] *– 'Possession of prohibited*

[108] http://www.iwf.org.uk/ (21 Oct 2009).
[109] The IWF remit also covers obscene content and racial hatred.
[110] The exact design of CleanFeed is not in the public domain.
[111] http://www.publications.parliament.uk/pa/cm200809/cmbills/072/09072.25-31.html#j3_100a (accessed on 01 June 2010).

images of children' (real child pornography is prohibited under the *Protection of Children Act*[112]).

Research indicates that the spread of conventional pornography (featuring no children) and easy access to it, especially via the Internet has led to a reduction (or at least no increase) in rape, sexual assault and sexual abuse [Kendall, T. 2007] [Diamond, M. 2009] as it allows the user to satisfy their fantasies and provides a cathartic release.

During the consultation period of the *Coroners and Justice Bill* the Home Office and Police expressed disgust at some manga (Japanese comics), but cited only a single case in which police raided an individual and found them in possession of only cartoon images [Ozimek, J., 2009, 17 March]. There is no consensus regarding the 'enabling' properties of child pornography, some experts follow the generally accepted view [Carr, J. 2003], other argue that it is analogous to conventional pornography, and as such computer generated images may even help prevent child abuse [Lazarová, D. 2010].

Crisp Thinking[113] has developed a tool described as an *Anti-Grooming Engine* (AGE) which builds a behavioural profile of a child's chat using Bayesian inference [Crisp Thinking, 2008]. *Crisp Thinking* claim 98.4% accuracy in independent tests [M2 Presswire. 2008], but beyond this headline figure, results, test data and methodology are not in the public domain[114]. While the range of products offered by *Crisp Thinking* are probably useful for helping to protecting younger children and blatant grooming attempts, their value in protecting teenagers who are discussing sex within their peer group are more questionable. While the algorithms and test methodology remain secret, the efficacy must be treated with scepticism.

Perverted-Justice[115] is a US based organisation that exposes those who attempt to groom children in Internet chat rooms. They have a pool of volunteers who setup 'honey pot' profiles in chat rooms and wait to be approached. The organisation is mired in controversy - California Judge, Dallas Holmes described the *Perverted-Justice* witness "odd", "weird" and "repulsive" [Stokley, S., 2007], following the collapse of a trial. Former *Dateline NBC* (*Dateline* worked with *Perverted-Justice* to create the US TV show *To Catch a Predator*) producer Marsha Bartel alleged *Perverted-Justice* begged individuals to come to sting locations [Franklin, K., 2007]. The activities of *Perverted-Justice* are of such concern to some that they have led to the formation of *Corrupted-Justice*[116] an organisation that researches and exposes the tactics of *Perverted-Justice* and works with attorneys and law enforcement.

Mitchell et al [Mitchell, K., Wolak, J., & Finkelhor, D., 2005] found that police stings (rather than vigilante/media operations) generally matched genuine 'real life' grooming and result in high conviction rates. They feel these proactive investigations are beneficial and encourage policy makers to support them. dana boyd [sic] questions the value of these activities, noting that if you seek solicitation you will find it - most teens ignored any solicitation and were not bothered, considering it a 'spam like' annoyance [Lordan, T., Finkelhor, D., Ybarra, M., Lenhart, A., boyd, D., 2007]. Those children who are groomed are mostly sexual mature teenagers and willing (if misguided) participants in their grooming.

DIFFICULTIES POSED TO RESEARCHERS

The more mundane areas of online child protection such as the effectiveness of filtering can be tested using standard laboratory methods. People however, are the most significant

[112] http://www.statutelaw.gov.uk/content.aspx?parentActiveTextDocId=1502057&ActiveTextDocId=1502059 (accessed on 01 June 2010).
[113] http://www.crispthinking.com (10 Nov 2009).
[114] The authors have asked to see information.
[115] http://www.perverted-justice.com/ (10 Nov 2009).
[116] http://corrupted-justice.com/about.html (11 Nov 09).

component of this system and the most unreliable component of any system. To this extent an examination of the human factors and stakeholders is of primary importance.

In order to gain a proper understanding of the problem it is essential to engage in stakeholder analysis, the most visible stakeholders in online child protection being the children, their parents, child abusers and service providers. Suitable access to two of these groups – children and child abusers – poses a challenge.

Typically stakeholder analysis might include interviews, questionnaires and observing the stakeholders. The very nature of the subject means that some of these questions would be of a sexual and personal nature. For prepubescent children this is unethically and possibly illegal as it risks exposing them to sexual information beyond their understanding. It is likely post pubescent children will be sexual aware and better informed, however their parents may still object to them taking part.

Child abusers and paedophiles are not going to participate unless they are already in the criminal justice system – as, to participate would likely involve admission of illegal activities. Young [Young, K., 2005] engaged 22 child sex offenders in forensic interviews. While her findings are not in dispute, the sample size is small, and questions must be asked about the motivations of those who took part. Those in the criminal justice system may lie during questioning, possibly being in denial, attempt to minimise or mitigate their guilt, or blame-shift [Lanning, K., 2001]. Further, they may feint cooperation in order to curry favour with prison authorities. Consequently this reduces the value of any information obtained from them.

The tabloid press reports paedophiles and child abusers with loathing and disgust [Moore, J., 2008], and is only interested in regarding them as monsters that require punishment. A proper systems evaluation demands stepping back from knee-jerk rhetoric and 'comfortable' stereotypes to consider what factors maybe be influencing child abusers. Could the behaviour, albeit misguided, of children be a contributory factor? Experts [Lanning, K., 2001] have noted that society does not want to believe such things are possible. Whether tabloids drive or reflect public opinion - child abusers are a despised group. Some controversial subjects such as recreational drug use may find advocates among societal commentators, academics, or even a subset of society at large. Child abusers and paedophiles have no public advocacy[117].

Researchers who engage in any questioning of this status quo risk being labelled as unethical and playing into the hands of paedophiles. They may find colleges and organisations are unwilling to work with them for fear of reduced funding. One university even refused to publish a doctoral thesis. Much of this is believed to stem to fear of how an individual or organisation will be treated by the tabloid press [Newman, M., 2009].

With academics and organisations working in fear of the tabloids, negative PR and public reaction, how can these difficult and sensitive subjects be properly examined?

There are some researchers conducting serious research in the problem space including the *Crimes Against Children Research Center*[118]and danah boyd[119] [sic]. Indeed their work on the *Internet Safety Technical Taskforce* (ISTTF) [Internet Safety Technical Taskforce, 2008] has

[117] Groups such as North American Man/Boy Love Association (NAMBLA) and individuals such as Tom O'Carroll are advocates for child sexual liberation, but these are marginal and ostracised, with little or no public platform in the mainstream media.
[118] http://www.unh.edu/ccrc/index.html (accessed 11 Nov 2009).
[119] http://www.danah.org/ (accessed 11 Nov 2009).

drawn much public criticism from the Attorneys General[120] (AG) of North Carolina and Pennsylvania [Jones, S., 2009]. The AG have in turn had their argument analysed and been thoroughly rebuked for political grandstanding and pandering to public fears [boyd, d., 2009], [Willard, N., 2009].

A system oriented meta-analysis is necessary to gain a full picture where the threat of negative PR hangs over researches. Government and law enforcement statistics, adolescent development, medical and psychological journals, parliamentary questions and testimony, and social commentary and opinion all provide sources that can be valuable. Researchers should always approach information with a critical eye, but this is especially important with government data, which maybe (possibly for political reasons) improperly reported, analysed or subject to withheld methodologies [Oates, J., 2009, 22 Sept], [Oates, J., 2009. 30 Sept].

DISCUSSION

It is the opinion of the authors that online child protection is a poorly understood problem space. Industry is offering products, services and advice to protect children, these offer some degree of protection for prepubescent children who need shielding from accidental exposure and there own naivety. Older children who are actively seeking contact with strangers (in social networks and chat rooms) and discussing sex and relationships are entirely misjudged either through misunderstand or a wilful disregard of their behaviour in the 'real world'.

In order to deliver real improvements to child safety and not security theatre the industry must first understand the problem space. This understanding begins with the child. The *Crimes Against Children Research Center* conducted the Youth Internet Safety Survey (YISS-1) [Crimes Against Children Research Center, 2009] in 2000 and repeated it in 2005 (YISS-2). This was a nationally (US) representative sample of behaviour and victimisation of 10-17 year olds on the Internet. With the rapidly changing nature of the Internet these surveys are becoming dated (*Facebook* was launched in 2004). Although the Internet is a global medium, do culture differences affect the risk i.e. would the results of the UK survey match those of the US? The UK Council for Child Internet Safety (UKCCIS)[121] a cross sector organisation investigating child safety online are investigating a survey of this type.

The tired mantras about stranger danger and not divulging personal details are unrealistic and widely ignored. With this in mind what recommendations can be made to children? Wolak et al [Wolak, J., Finkelhor, D., & Mitchell, K., 2004] recommend a programme of education and awareness e.g. making children aware that those who seek sexual relationships with them probably do not have their best interests at heart. It must be recognised that adolescents will break some rules, they must be engaged with, and informed of the risks and dangers, they must also feel they can turn to parents (teachers and other responsible adults) if they do find themselves embroiled in solicitation, bullying or other harmful activity. Parents etcetera, must accept boundary testing and rule breaking are normal, and that their children will be involved in things they disapprove of. They must also understand that if the child is too scared to come forward under such circumstances things will likely get worse.

Parents should continue to be vigilant for risk factors such as abuse, isolation, bullying, substance abuse, and self esteem issues etc, and take intervening action if necessary. It must also be recognised that some parents are negligent or ineffective. Safe Internet behaviour should be included in the education curriculum – tailored to the child's age and sexual maturity. In the UK teachers are already trained to spot potentially troubling behaviours and have procedures in place to intervene where necessary. This training should be standardised and extended to include an understanding of Internet technologies and how these threats

[120] The position of a state Attorney General is political, the majority of Attorneys General are elected included all mention here.
[121] http://www.dcsf.gov.uk/ukccis/index.cfm (accessed 12 Nov 2009).

(which are essentially the same as those children face offline) manifest themselves. Internet threats do not present the same immediate threat as those in the real world (a child is not at risk of physical harm unless they go and meet somebody) and the child should be encouraged to log off and inform an adult if they feel threatened. Today's children are more 'connected' than ever before and this trend shows no sign of abating, as such adults should not dismiss a child's complaint of online harassment with advice along the lines of "well turn the computer off and do something else". This maybe adequate in the short term while issues are investigated, but in the mid- to longer term may have a detrimental effect on the child's development and peer relationships. Communicating with strangers should be treated with caution but not proscribed. Parents and children should work together to devise rules on Internet communication. A strong parent-child relationship, where children understand the risks and are comfortable discussing them with parents, not authoritarian rules or technology will provide the greatest protection to children.

The goal of the media is to maximise viewers/listeners/readers, a good way of doing this is sensational, mawkish, scare stories. Although it is highly implausible the media should be discouraged from reporting child protection in this fashion.

Laws need to be in place to protect children but these laws must be based on evidence not subjective morality or a desire to appease, especially when there exists the very real risk such law might actually be harmful.

CONCLUSION

This paper has explored online child protection from a systems perspective, describing how it became such a prominent subject and how its sensitivities pose challenges to research. Using systems thinking it has shown that conventional wisdom is misinformed about the dangers – sharing personal information on the Internet does not, in itself increase risk, and that some known risky activities such as discussing sex in chat rooms, have potentially unseen, beneficial side effects. Children, especially adolescents, the key stakeholder in the system, are poorly understood, advice given to them is outdated, unrealistic and likely to be ignored. Technology has a role to play particularly protecting younger children against accidental exposure to unwholesome content. Filters can block access to undesirable material but the quality and reliability of filtering technology is limited. Here more aggressive blocking, with a higher false positive rate (over blocking) maybe acceptable, but the best protection for younger children is close supervision.

The role of technology in protecting technological au fait, head strong teenagers is smaller. Even those without the technical proficiency to circumvent protection measures are still able to use a friend's computer or public Internet terminal. Continuing to ignore the behaviour of teenagers and develop products, services and advice in absence of what is known about them will see under performing solutions and ill informed users – given either a false sense of security or inaccurate understanding of the real risks. Parents are probably most fearful of younger children being tricked into meeting, and then forcibly molested by paedophiles. Research indicates this scenario is extremely rare. Victims are overwhelmingly sexually mature teenagers who are seduced by attention and affection, and are aware their seducer is an older man. Adolescents and parents alike need to be educated and informed of real dangers, not mawkish, scare-mongering anecdotes from the media and other sources.

Real life abuse, bullying, peer group isolation, substance and alcohol abuse, self esteem and confidence problems all increase the risk of being victimised on Internet. It is the authors' belief that the foundation of child safety on the Internet is different for prepubescent and post pubescent children; prepubescent children are best protected with close supervision and oversight. The best way parents can protect post pubescent adolescents is by forming a close relationship with them, working with them to define rules, rather than dictating to them.

Parents should expect their child to, at least occasionally be ‘up to no good’, but with a strong trusting relationship in place the child will turn to the parent whenever they are upset or in trouble.

REFERENCES

American Psychiatric Association. (2003). *Medical library | medem.com.* Retrieved 06/06/2009, 2009, from http://www.medem.com/medlib/article/ZZZUZRUZGLC

BBC (2004). *BBC NEWS | magazine | teenage kissing: The new sex crime?* Retrieved 15/12/2009, 2009, from http://news.bbc.co.uk/1/hi/magazine/3672591.stm

boyd, d. (2009). *Apophenia: Doing the math on MySpace and registered sex offenders.* Retrieved 11/08/2009, 2009, from http://www.zephoria.org/thoughts/archives/2009/02/06/doing_the_math.html

Carr, J. (2003). Child abuse, child pornography and the internet No. 1). London: NCH.

Cellan-Jones, R. (2009). BBC - dot.life: Can we block child abuse sites? Retrieved 16/06/2009, 2009, from http://www.bbc.co.uk/blogs/technology/2009/02/can_we_block_child_abuse_sites.html

Clayton, R., (2005) Failures in a hybrid content blocking system. Cambridge: University of Cambridge.

Crimes Against Children Research Center (2009). Retrieved 30/10/2009, from http://www.unh.edu/ccrc/youth_internet_safety_survey.html

Crisp Thinking (2008). CRISP THINKING'S ANTI-GROOMING TECHNOLOGY ACHIEVES 98.4% ACCURACY IN INDEPENDENT TESTING [press release].

Diamond, M. (2009). Pornography, public acceptance and sex related crime: A review. International Journal of Law and Psychiatry, 32(5), 304-314.

Franklin, K. (2007). American chronicle | predator show slammed. Retrieved 30/04/2009, 2009, from http://www.americanchronicle.com/articles/view/29223

Internet Safety Technical Taskforce (2008). *Enhancing child safety and online technologies* No. 1). Harvard University: The Berkman Center for Internet and Society. Retrieved from http://cyber.law.harvard.edu/pubrelease/isttf/

Jones, S. (2009). *MySpace: 90,000 sex offenders removed from site | technology | guardian.co.uk.* Retrieved 07/08/2009, 2009, from http://www.guardian.co.uk/technology/2009/feb/04/myspace-social-networking-sex-offenders

Kendall, T. (2007). Pornography, rape, and the internet. Clemson, South Carolina: Clemson University.

Lanning, K. (2001). Child molesters: A behavioral analysis (Fourth Edition) National Centre for Missing & Exploited Children.

Lazarová, D. (9th Feb). Child porn consumers safe from prosecution in the czech republic - radio prague. Retrieved 29/07/2010, 2010, from http://www.radio.cz/en/article/88189

Livingstone, S., & Helsper, E. (2007). Taking risks when communicating on the internet: The role of offline social-psychological factors in young people's vulnerability to online risks. *Information, Communication & Society, 10*(5), 619-644.

Lordan, T., Finkelhor, D., Ybarra, M., Lenhart, A., boyd , D., (2007) Internet caucus advisory committee [panel discussion].

McCarthy, C. (2009). *Whee! new numbers on social network usage | webware - CNET*. Retrieved 08/07/2009, 2009, from http://news.cnet.com/8301-17939_109-10160850-2.html

Mitchell, K., Finkelhor, D., & Wolak, J. (2001). Risk factors for and impact of online sexual solicitation of youth. *Journal of American Medical Association, 285*(23), 3011-3014.

Mitchell, K., Wolak, J., & Finkelhor, D. (2005). Police posing as juveniles online to catch sex offenders: Is it working? *Journal Sexual Abuse: A Journal of Research and Treatment, 17*(3), 241-267.

Moore, J. (2008). *Jane moore | make these vile monsters suffer | the sun |News|Columnists|Jane moore*. Retrieved 28/05/2009, 2009, from http://www.thesun.co.uk/sol/homepage/news/columnists/jane_moore/875869/Jane-Moore-Make-these-vile-monsters-suffer.html

National Research Council (2002). Technical, business, and legal dimensions of protecting children from pornography on the internet. *Committee to Study Tools and Strategies for Protecting Kids from Pornography and their Applicability to Other Inappropriate Internet Content*, 1-145.

Newman, M. (2009). *Times higher education - paedophilia research riles and titillates the academy*. Retrieved 12/10/2009, 2009, from http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=408084&c=2

Oates, J. (2009, 22 Sept). *Home office stonewalls ID findings • the register*. Retrieved 15/10/2009, 2009, from http://www.theregister.co.uk/2009/09/22/id_card_research_blanked/

Oates, J. (2009. 30 Sept). *Home office declines to detail DNA-for-foreigns trial • the register*. Retrieved 15/10/2009, 2009, from http://www.theregister.co.uk/2009/09/30/dna_aslyum/

Ozimek, J. (2009). *IWF chief: We don't need crusaders • the register*. Retrieved 10/11/2009, 2009, from http://www.theregister.co.uk/2009/09/08/iwf_peter_robbins_interview/page2.html

Ozimek, J. (2008, 14 July). *Criminal record checks could hit over 14 million people | the register*. Retrieved 20/08/2008, 2008, from http://www.theregister.co.uk/2008/07/14/crb_checks_total_analysis/

Ozimek, J. (2008, 9 October). Porn, abuse, depravity - and how they plan to stop it • the register. Retrieved 25/02/2009, 2009, from http://www.theregister.co.uk/2008/10/09/policing_internet_one/

Ozimek, J. (2009, 17 March). Govt uses obscenity law to stuff up cartoon sex loophole • the register. Retrieved 28/01/2010, 2010, from http://www.theregister.co.uk/2009/01/19/evil_cartoon_badness/

Smith, A. (2007). Teens and online stranger contact Pew Internet & American Life Project.

Stokley, S. (2007). *'To catch a predator' sex stings net mixed results | inland news | PE.com | southern california news | news for inland southern california*. Retrieved 14/08/2008, 2008, from http://www.pe.com/localnews/inland/stories/PE_News_Local_R_dateline28.6b3814.html

Subrahmanyam, K., Greenfield, P. M., & Tynes, B. (2004). Constructing sexuality and identity in online teen chat room. *Applied Developmental Psychology, 25*, 651-666.

TheyWorkForYou.com (2009). Coroners and justice bill: 3 feb 2009: Public bill committees (TheyWorkForYou.com): 26/03/2009 Retrieved from http://www.theyworkforyou.com/pbc/2008-09/Coroners_and_Justice_Bill/02-0_2009-02-03a.1.0?s=speaker:11480#g1.278

Tromovitch, P. (2009). Manufacturing mental disorder by pathologizing erotic age orientation: A comment on blanchard et al. (2008) [Letter to the editor]. *Archives of Sexual Behavior, 38*(3)

Willard, N. (2009). *Why age and identity verification will not work - and is a really bad idea* Center for Safe and Responsible Internet Use.

Wolak, J., Finkelhor, D., & Mitchell, K. (2004). Internet-initiated sex crimes against minors: Implications for prevention based on findings from a national study. *Journal of Adolescent Health, 35*(5), 424.e11-424.e20.

Wolak, J., Finkelhor, D., Mitchell, K., & Ybarra, M. (2008). Online "predators" and their victims myth, realities and implications for prevention and treatment. *American Psychologist, 63*(2), 111-128.

Wolak, J., Mitchell, K., & Finkelhor, D. (2003). THE EXPOSURE OF YOUTH TO UNWANTED SEXUAL MATERIAL ON THE INTERNET A nation survey of risk, impact and prevention. *Youth & Society, 34*(3), 330-358.

Ybarra, M., & Mitchell, K. (2004). Online aggressor/target, aggressors, and targets: A comparison of associated youth characteristics. *Journal of Child Psychology and Psychiatry, 45*(7), 1308-1316.

Young, K. (2005). Profiling online sex offenders, cyber-predators and pedophiles. *Journal of Behavioural Profiling, 5*(1)

6. CONCLUSIONS AND FURTHER WORK

This section will examine what has been learned from each of the pieces of work discussed in the main body, the author's contribution to the body of knowledge, publications and reports and suggestions for furthering the state of the art.

6.1 Identity Management

More and more people are using the Internet for more and more services. Typically these services require a user to register and issue the user with login credentials – a username and password. This has two side effects – the users are saddled with a plethora of usernames and passwords and are regularly disclosing personal information to service providers. Faced with ever more login credentials users often succumb to password fatigue, employing insecure practices such as reusing credentials across multiple websites and services and writing down login credentials in clear text. Disclosing personal details to so many providers' leaves users open to the nuisance of direct marketing campaigns and spam, as well as the risk of identity theft.

These problems stem from the underlying nature of the Internet – it was never designed to support an identity system. Additional systems have been developed to support elements of identity, Public Key Infrastructure (PKI) for example is used for email encryption and SSL, usernames, passwords and sometimes hardware devices are used to authenticate users. *Microsoft Passport*, a single sign-on web commerce service, was the most notable attempt to address the problem but failed for reasons of privacy, trust and consumer buy-in.

Federated identity management seeks to address this by having the user register with an identity provider. The identity provider would then attest the user's identity to relying parties – websites and other service providers who would previously have required the user to register. This paradigm also places the user in control of their personal information; the user decides how much personal information the identity provider may reveal to the relying party. If the relying party demands too much personal information the user has the option to cancel and withdraw from the process.

In order to be successful a new identity metasystem needs to provide federation, login security, information security, user centricity and standardisation.

Technically none of these elements are insurmountable; indeed standards and protocols exist to build a federated identity infrastructure. This has led to the adoption of identity management within the enterprise (user centricity being delegated to the organisation).

The author has shown that whilst there exists no major technical obstacles, without a breakdown in Internet trust and security, the success of a consumer federated identity metasystem is unlikely. The enterprise is happy to pay for identity management – it simplifies user management and the enterprise is already liable for its data. The cost is offset by administrate savings. No business model exists in the consumer space. Consumers like functionality but are less interested in security and are unwilling to pay for services given a choice. Service providers (relying parties) will be resistant to losing the valuable marketing opportunities provided in personal details and sales tracking. Identity providers will face liabilities in the event of data loss and fraud but have no revenue stream. This will prevent an identity provider moving beyond the low value applications such as blogging that are currently supported.

If the Internet were to see a breakdown of trust and security with a substantial percentage of transactions being fraudulent and leading to identity theft this could be the driver required to see a federated identity metasystem, perhaps being championed by the government or financial sector. Until such time the status quo will persist with a hardening of existing methods e.g. more complex password rules and increased use of OTPs. The author has published these findings on the future of identity and the J-curve.

While there is undoubted progress to be made in simplifying the process of federation and forming dynamic trusts, the big challenges are overcoming user apathy and developing a sustainable business model. The author feels this is dependent on the collapse of Internet trust and security, or a major industry and government push.

6.2 Text Analysis

Education and recreation are seeing more children spending increasing amounts of time online. This frequently involves interaction with individuals neither they, nor their parents have met in the 'real world'. The absence of this real world interaction makes it much more difficult to determine the identities of those individuals with whom children come in contact. The Internet offers a great deal of anonymity between users.

The vast majority of people on the Internet, as in real life, present no harm to children; their contact will be beneficial or at least benign. A small minority of people do however pose a threat to children and the anonymity of the Internet reduces their risk of detection.

The author has investigated the feasibility of detecting child grooming based on the content of Internet chat – interactions in email, chat rooms, IM, social networks and message boards. While semantics can be effective in the analysis of language, its efficacy is greatly reduced when language does not adhere to normal definitions i.e. where spelling, grammar and syntax do not follow rules. Informal lingo, slang and the deliberate use of misspelling and poor grammar (sometimes called 'text speak' and/or 'leet speak') are a common feature of youth conversation. A statistical method can overcome the limits of the linguistic method, it has no 'understanding' of language, instead it analyses patterns.

Statistical models need numerical data; this is generated by feature extraction - assigning numerical values to language characteristics. Trigram analysis for example is based on the frequency of trigram occurrence.

Bayesian trigram analysis has been shown to be the most accurate classifier given the linguistic nature of online chat among children. The author attempted to establish if this method could be improved upon using a novel feature extraction method consisting of token vectors made multiple coefficients (multiple character strings) and Gaussian Mixture Models.

Around 80% of grooming is obvious with groomers being honest about their age and sexual intention. To test the classifier the author selected four texts *Nineteen Eighty-Four, The War of the Worlds, Fahrenheit 451* and *On the Origin of Species*. The first three books are broadly similar works of fiction, the latter a scholarly work of biology. In experiments *On the Origin of Species* was used to represent the blatant grooming, the assumption being this blatant grooming would be substantially different from regular chat as *On the Origin of Species* is substantially different from the other works.

Assuming parents and children could easily spot the blatant grooming attempts, the real challenge is differentiating subtle grooming from normal chat. This challenge is essentially the same as successfully classifying the three fictional works.

Experimental results showed the novel feature extraction/GMM method proved very unreliable and substantially inferior to Bayesian trigram analysis.

The novel feature extraction/GMM process was computationally expensive and unreliable, with statistical expertise the process could be improved but the author doubts it would become the leading classifier of undisciplined Internet conversations. 80% of grooming is blatant; the other subtler grooming is known to involve building a relationship by showing interest and affection. Showing interest and affection – dating, is normal teenage behaviour. Differentiation is very difficult if not impossible because the conversation content is essentially the same. Further work in this area is possible but the author feels it would not represent the best use of resources in the child protection ecosystem.

6.3 Steganographed Custom Emoticons

Instant messaging is one of the most popular online activities among children. The *.NET Messenger Service* often simply *MSN* is a popular IM platform with over 80 million users in Europe.

Emoticons are key sequences, images and animations that are used to add information content to text by expressing emotions that would be present in non-verbal communication if the conversation were taking place face-to-face. Simple ASCII

emoticons began to take-off in the early 1980s with the suggestion of using :) to express a joke or being happy. With the rise of rich media these have become small images and even animations. Many applications now intercept the keystrokes and display rich media emoticons. As well as having a default set of emoticons the *.NET Messenger Service* also supports the use of user generated custom emoticons.

Steganography is the art of hiding information in plain sight, for example hidden writing with invisible ink. It is possible to steganograph digital images in a number of ways e.g. using the least significant bits of a noisy file.

Using steganography techniques it is possible to make an ordinary image unique by changing the aRGB (alpha, red, green, blue) channels of a pixel. This subtle change is imperceptible to the human eye but readily detected with hashing functions.

The author developed software to modify the aRGB values of a single pixel of a common emoticon to generate 625 emoticons that appeared identical to the human eye but were in fact unique. These emoticons can be loaded into any *.NET Messenger Service* client (the service is run on open protocols) that supports custom emoticons.

Clients on the *.NET Messenger Service* do not transmit the emoticon images associated with default emoticons, only the key sequence is sent. Both the transmitting and receiving client show their local copy of the emoticon. Custom emoticons do not already exist at the receiver and so must be included with the key sequence. Custom emoticons are buried deep within the receiver's file system – retrieving them is a non-trivial exercise. The emoticon transfer is handled as a drive-by download, i.e. it is automated background activity requiring no action on the part of the receiver.

A 'steganographed' custom emoticon can be used as a type of 'digital fingerprint' left on the computer of anyone who chats with a child. This can later be used by authorities as evidence that communications have taken place perhaps in a case of suspected child grooming. If the user saves and re-uses the custom emoticon, the act of loading it in the application changes the hash, meaning any attempt to imitate the original child (user) or spoof the emoticon is easily detected.

Internet criminals are detected using one of three methods, identification via IP address, disclosure of personal information that leads to real world identification, or they come to the attention of law enforcement by other methods. The hierarchical governed nature of public IP addresses (necessary for Internet use) means an individual can be easily traced to an ISP. UK ISPs are required to retain subscriber records for 12 months.

Evading IP address identification is a trivial process that can be accomplished through numerous ways – proxies, onion routing, 'anonymising' services, 'wardriving', use of WiFi hotspots etc. Despite this simplicity many Internet criminals are caught through IP address identification. In the same vein, though it is possible to detect and delete the custom emoticons, or use a public terminal at a library or Internet café, many individuals will not bother, thereby preserving the unique custom emoticon on their computer.

The system could be implemented in a number of ways. The author feels the best way would be to operate the system as a managed service. Parents would register with the service provider and download a custom client for the *.NET Messenger Service*. This client would contain the custom emoticon, the details of which would be recorded alongside the registration details held at the service provider. The custom emoticon could be automatically 'pushed out' (without the sender needing to manually invoke the custom emoticon) each time the child begins chatting with a new contact. Any emoticons recovered by authority could be compared against the service provider database.

The concept has been shown to work and the technologies (digital imaging, hashing and the *.NET Messenger Service*) are mature. Further work developing the custom client and service could be a cheap and simple activity building on existing open source client.

BT considered patenting the process, but instead decided to publish (prevent patentability) and allow anyone to develop the idea without royalty restrictions.

6.4 Online Child Protection

During the author's technical work relating to Internet child protection it became apparent that technological defences were only a small part of large complex issue. A problem ideally suited to systems analysis.

The online child protection ecosystem contains much fear, uncertainty and doubt. Politicians, products and services providers and the media have 'muddied the waters' with alarmist anecdotes, selective reasoning and worst case scenarios. This combined with the sheer volume of information on the subject has left children, parents and society poorly informed about the nature of the issues at large.

Systems theory teaches that no part of a complex system can be considered in a vacuum, the interfaces and interactions between subsystems and components must be taken into account. Human factors engineering reveals that people are the unreliable component in any mature system. Human behaviour and culture can easily upset strategy and development [111]. Systems developed without an appreciation of stakeholder requirements rarely achieve optimum performance. In order to understand and assess the problem, proper requirements capture and stakeholder analysis is essential. Many of the products, services and even advice around child Internet safety ignore the behaviour of children.

Perhaps the single most significant consideration is the difference between prepubescent and post pubescent children. Prepubescent children are easier to protect – while they are more naïve they are naturally more dependent on their parents and readily accept close supervision and oversight. Post pubescent children, teenagers, are a more difficult demographic, at significantly more risk. Physically mature and sexually aware, they yearn for adult levels of independence and privacy, but lack psychological maturity and emotional control. They are likely to be headstrong, defiant, secretive and testing/dismissive of authority – these are normal developmental behaviours. Technically au fait children may even have the ability to circumvent any technological measures; others may bypass them by using a friend's computer or public terminal.

Children face three primary risks online, solicitation, exposure to unwanted content and bullying. Solicitation typically comes in the form of a crude comment in a conversation; the vast majority of children adeptly handled the incident and suffered no negative effects from it. Unwanted content is not just limited to pornography; it can also include violent content, bigotry and the promotion of illegal activities such as criminal skills and drug use. As with solicitation the majority of children were unbothered by exposure to unwanted content. It is not uncommon for teenage boys in particular to actively seek this type of content. Cyberbullying is a poorly understood, poorly defined problem. It is thought a significant minority of children are involved either as bully, victim or both. The effects, both immediately and in the long term, may be more severe than real world bullying. With ever more devices connected and always online, combined with more peer group interaction online, cyberbullying is increasingly difficult to escape. Victims of bullying are also at greater risk of other forms of Internet victimisation.

There is no such thing as a typical child abuser, their motivations, methods, socioeconomic status etc can't be discreetly categorised, they fall along a continuum. The vast majority of child abuse takes place within the family or at the hands of someone known to the child. Child Internet groomers are typically honest and open about their ages and sexual intent; the majority of victims who meet groomers in real life know they are meeting an older man for sex. Grooming typically revolves around showing the victim affection and taking an interest in them. Abduction and forcible rape are extremely rare. Victims often refuse to help authorities, citing strong emotional feelings towards their abuser. Children who are already abused, neglected, bullied, depressed, suffering peer group isolation or otherwise psychologically troubled are at greater risk of becoming the victims of Internet initiated sex crimes. They are often active, even enthusiastic participants in their own abuse.

The government needs to be clear about the dangers online and back a concise safety message grounded in the realistic known behaviours of children. *ThinkUKnow* is by and large an excellent source of information, but risks being undermined by CEOP's and the government's media agenda. Good advice and good legislation is based on evidence, not untested ideas, anecdotes and subjective morality. Posting personal information and photos on the public Internet may not be a good idea, but there is no

evidence that child abusers are ‘data mining’ this information in the search for victims. People may find computer generated (artificial) child pornography repugnant but there is the possibility that it makes children safer by providing paedophiles catharsis.

Education provides the best, most cost effective way to protect children on the Internet. *ThinkUKnow* is generally an excellent resource with age appropriate material but its’ message needs to be pushed out through schools perhaps featuring role play to help guide children in case they are solicited, bullied or exposed to unwanted content. Education should not be limited to children; parents cannot be dismissive or fearful of technology because just they don’t understand it. A simple, high level overview need not be technical but would help the parents establish a common ground and engage in dialogue with their children. It will help them to understand the environment, the appeal, the benefits and the risk. Such education should also be mandatory for teachers and social workers.

A small but not insignificant number of children are bothered by solicitation and unwanted exposure to harmful content, this is more likely if the solicitation is persistent and aggressive and the content is explicit and depicts pain. This may be exacerbated if the child is going through puberty, becoming sexually aware but not fully able to comprehend the situation. In addition to covering such scenarios in education, the child needs a strong, supportive and understanding parental relationship.

Some degree of boundary testing and ‘getting up to no good’ is to be expected from teenagers, however if they become involved in a situation they are unable to handle they must feel able to turn to a parent or other responsible adult. In the absence of this they may worsen the situation because of feelings of isolation or fear of excessive punishment.

Some children will always suffer abuse, solicitation, bullying and exposure to unwanted content. Technology can provide a measure of protection, especially protecting younger more naïve children from exposure to harmful content. It is less beneficial at protecting older children. The child-parent relationship is important at all

age levels, younger children need supervision, older children need support and trust. The Internet is now a major factor in both the education and leisure of children, parents need to understand to the ‘basics’ in order to be properly involved in their child’s development. Parents would also benefit from properly understanding the risks involved in child Internet use and understand that alarming media anecdotes are news worthy because of their rarity.

Online child protection is an enormously broad subject covering technology, psychology, sociology and behaviour, legislation and civil liberties. Most people have an opinion and many hold strong beliefs. Internet child protection is never far from the headlines. Yet much of what people believe - the generally accepted view - is an inaccurate representation of the facts. The author has shown the subject is too complex to properly address solely at the subsystem level and that focusing on an individual variables at the expense of the whole is inefficient and retards progress.

Further work should involve thought leaders asking serious questions about security theatre and the real value of measures legal, technical and social used to protect children. A broad survey similar to YISS needs to be performed to understand how children in the UK are using the Internet and what dangers they have encountered. A syllabus which educates and engages children, but which also understands their behaviour, needs to be developed. Training complimentary to this syllabus needs to be developed for teachers, social workers and parents, not only covering the technology but realistically portraying the issues – the risks and how their children’s behaviour contributes to this.

Social scientists need to work to define common metrics so that research can be meaningfully compared. They also need to work with industry to show that while products and services meet technical requirements, they do not necessarily meet user requires or address user behaviour.

7. REFERENCES

The data used in this thesis represents the current best thinking. Some sources are older where the thinking is unchanged (e.g. human behaviour) or where little relevant data exists (e.g. cyberbullying).

7.1 Identity Management

[1] K. Cameron. (2011, Sun, 27 Mar). Broken laws of identity lead to system's destruction. Kim Cameron's Identity Weblog [Online]. Available: http://www.identityblog.com.

[2] J.P. Rangaswami. (2009, 5 February). Identity – confused of calcutta. [Online]. 2011(5/22/2011), Available: http://confusedofcalcutta.com/category/identity/.

[3] V. Bertocci, G. Serack and C. Baker. (2008, *Understanding Windows CardSpace: An Introduction to the Concepts and Challenges of Digital Identities (Independent Technology Guides)* Available: http://isbndb.com/d/book/understanding_windows_cardspace_an_introduction_to_the_conce.

[4] E. Harrington, "Identity management in the open group," in *SEWP Identity Management Symposium*, 2004.

[5] K. Finklea, "Identity theft: Trends and issues," Congressional Research Service, 27 May. 2009.

[6] K. Cameron. (2005, May 2005). The laws of identity. [[Online]]. *2007(21 May 2007),* pp. 14 (printed). Available: http://www.identityblog.com/stories/2004/12/09/thelaws.html

[7] G. Fish. BusinessWeek debate room employers, get outta my facebook. [Online]. *2009(10/02/2009),* pp. 1. Available: http://www.businessweek.com/debateroom/archives/2008/03/employers_get_o.html.

[8] R. Bennett. (25 March, 2008). Plea to ban employers trawling facebook - times online. [Online]. *2009(10/02/2009),* pp. 1. Available: http://technology.timesonline.co.uk/tol/news/tech_and_web/article3613896.ece.

[9] T. Martin, C. Durbin, M. Pawlewski and D. Parish. (2010), Future vision of identity. *IJLSE* [Online]. *3(1/2),* pp. 86-98. Available: http://www.inderscience.com/search/index.php?action=record&rec_id=31825&prevQuery=&ps=10&m=or.

[10] D. Goodin. (2007, 3rd December). TJX agrees to pay banks $41m to cover visa losses • the register. [online]. *2010(11/06/2010),* Available: http://www.theregister.co.uk/2007/12/03/tjx_settlement_agreement/.

[11] J. Oates. (2007, 20 November). Darling admits revenue loss of 25 million personal records • the register. [Online]. *2009(20/02/2009),* Available: http://www.theregister.co.uk/2007/11/20/hmrc_loses_lots_data/.

[12] I. Bremmer. (2006, September). *The J Curve: A New Way to Understand Why Nations Rise and Fall* [Online]. *2010(11/06/2010).* Available: http://www.jcurvebook.com/index.html.

[13] Anonymous (2005, 03 March). BBC NEWS | business | one in four 'touched' by ID fraud. [Online]. *2010(16/06/2010),* Available: http://news.bbc.co.uk/1/hi/business/4311693.stm.

[14] D. Chappell. Introducing windows CardSpace. [Online]. *2007(15/08/2007),* pp. 25 (Printed). Available: http://msdn2.microsoft.com/en-us/library/aa480189.aspx.

[15] J. Leyden. (2008, 3 December). Windows patching abysmaland getting worse • the register. [Online]. *2010(25/06/2010),* pp. 1. Available: http://www.theregister.co.uk/2008/12/03/secunia_patching_survey/.

[16] S. Granneman. (2004, 19 August). Infected in 20 minutes • the register. [Online]. *2010(25/06/2010),* pp. 1. Available: http://www.theregister.co.uk/2004/08/19/infected_in20_minutes/.

[17] Anonymous (2002, 8th August). BBC NEWS | technology | microsoft rapped over privacy failings. [Online]. *2010(01/09/2010),* pp. 1. Available: http://news.bbc.co.uk/1/hi/technology/2181508.stm.

[18] Anonymous (2003, 9th May). BBC NEWS | technology | flaw exposes microsoft ID service. [Online]. *2010(01/09/2010),* pp. 1. Available: http://news.bbc.co.uk/1/hi/technology/3013665.stm.

[19] P. Becker. (20006, 30 August). IdentityBlog - digital identity, privacyand the internet's missing identity layer. [Online]. *2010(02/09/2010),* pp. 1. Available: http://www.identityblog.com/?p=551.

[20] Facebook (not dated) Press room. [Online]. *2010(10/18/2010),* Available: http://www.facebook.com/press/info.php?statistics.

[21] D. Goodin. (2011, 10 May). Facebook caught exposing millions of user credentials • the register. [Online]. 2011(5/15/2011), Available: http://www.theregister.co.uk/2011/05/10/facebook_user_credentials_leaked/.

[22] D. Goodin. (2011, 10 May). Facebook caught exposing millions of user credentials • the register. [Online]. 2011(5/15/2011), Available: http://www.theregister.co.uk/2011/05/10/facebook_user_credentials_leaked/.

[23] J. Leyden. (2011, 21 January). Facebook defends security strategy • the register. [Online]. 2011(5/17/2011), Available: http://www.theregister.co.uk/2011/01/21/facebook_security_analysis/.

[24] B. Nussbaum. (2010, 24 May). Facebook's culture problem may be fatal - bruce nussbaum - the conversation - harvard business review. [Online]. 2011(5/18/2011), Available: http://blogs.hbr.org/cs/2010/05/facebooks_culture_problem_may.html.

[25] B. Acohido. (2010, 16 June). Facebook’s business model calls for wiping out privacy | the last watchdog. [Online]. 2011(5/18/2011), Available: http://lastwatchdog.com/facebooks-business-model-hinges-wiping-privacy/.

7.2 Text Analysis

[1] Internet Safety Technical Taskforce. (2008, 31 December 2008). Enhancing child safety and online technologies. The Berkman Center for Internet and Society, Harvard University. [Online]. Available: http://cyber.law.harvard.edu/pubrelease/isttf/

[2] Anonymous Europe surpasses north america in instant messenger users, comScore study reveals. [Online]. *2008(10/08/2008),* Available: http://www.comscore.com/press/release.asp?press=800

[3] Anonymous Statistics on canadian youth and IM. [Online]. *2008(10/08/2008),* Available: http://www.media-awareness.ca/english/resources/special_initiatives/wa_resources/wa_shared/backgrounders/statistics_youth_im.cfm

[4] N. Pendar, "Toward spotting the pedophile telling victim from predator in text chats," in *Semantic International Conference on Computing, 2007*, 2007, pp. 235-241.

[5] W. Ding, S. Yu, Q. Wang, J. Yu and Q. Guo. (2008, A novel naive bayesian text classifier. *Information Processing (ISIP), 2008 International Symposiums on* pp. 78-82.

[6] A. Schwartz, "Bayesian filtering," in *SpamAssassin* ,First ed., vol. 1, J. Gennick and D. Kelly, Eds. Sebastopol, California: O'Reilly, 2004, pp. 68-69-80.

[7] J. Provost. (1999, 1999). Naive-bayes vs. rule-learning in classification of email. University of Texas at Austin, Artificial Intelligence Lab, Austin, Texas. [Online]. Available: http://www.cnbc.cmu.edu/~jp/research/email.paper.pdf

[8] P. Graham. (2004, *Hackers & Painters: Big Ideas from the Computer Age.* Available: http://isbndb.com/d/book/hackers_painters

[9] Wikipedia contributors. (2011, 8 April). Bayesian spam filtering. [Online]. 2011Available: http://en.wikipedia.org/wiki/Bayesian_spam_filtering.

[10] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE,* vol. 77, pp. 257-286, February. 1989.

[11] J. Dugelay, C. Junqua, C. Kotropoulos, R. Kuhn, F. Perronnin and I. Pitas, "RECENT ADVANCES IN BIOMETRIC PERSON AUTHENTICATION," Eurecom, eurecom.fr

7.3 Steganographed Custom Emoticons

[1] M. Ybarra and K. Mitchell, "Online aggressor/target, aggressorsand targets: a comparison of associated youth characteristics," *Journal of Child Psychology and Psychiatry,* vol. 45, pp. 1308-1316, 2004.

[2] M. L. Ybarra and K. J. Mitchell. (2008, Feb). How risky are social networking sites? A comparison of places online where youth sexual solicitation and harassment occurs. *Pediatrics 121(2),* pp. e350-7.

[3] J. Hancock, C. Landrigan and C. Silver, "Expressing emotion in text-based communication," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* San Jose, Calif, 2007, pp. 929-932.

[4] O. Kharif. (2001, 2 April 2001). BW online | april 23, 2001 | the man who brought a :-) to your screen. [Online]. *2008(19/08/2008),* Available: http://www.businessweek.com/bwdaily/dnflash/apr2001/nf20010423_785.htm.

[5] Anonymous (2001, 14 December). Anti-terrorism, crime and security act 2001 (c. 24). [Online]. *2010(01/07/2010),* pp. 20. Available: http://www.opsi.gov.uk/acts/acts2001/ukpga_20010024_en_11.

[6] Anonymous (1999, 18 August 1999). Microsoft to publish MSN messenger service protocol to industry: Furthers commitment to instant messaging interoperability; provides solution to meet strong consumer demand for open

communications. [Online]. *2008(23/09/2008),* Available: http://www.microsoft.com/presspass/press/1999/aug99/protocolpr.mspx.

[7] R. Movva and W. Lai. (1999, August 1999). Instant messaging and presence protocol. [Online]. *2008(23/09/2008),* Available: http://www.hypothetic.org/docs/msn/ietf_draft.txt.

[8] Anonymous Europe surpasses north america in instant messenger users, comScore study reveals. [Online]. *2008(10/08/2008),* Available: http://www.comscore.com/press/release.asp?press=800.

[9] F. Furedi. (2001, 25 April 2001). Frank furedi: Making sense of parental paranoia. [Online]. *2008(02/07/2008),* Available: http://www.frankfuredi.com/index.php/site/article/112/.

[10] Ketchum Global Research Network, "Parents' internet monitoring study," National Center for Missing & Exploited Children and Cox Communications, 2005.

[11] M. L. Ybarra and K. J. Mitchell. (2008, Feb). How risky are social networking sites? A comparison of places online where youth sexual solicitation and harassment occurs. *Pediatrics 121(2),* pp. e350-7.

[12] Anonymous 4chanarchive - /b/ - brb church - chris forcand. [Online]. *2010(03/09/2010),* Available: http://4chanarchive.org/brchive/dspl_thread.php5?thread_id=42828652&x=brb+church+-+chris+forcand.

[13] J. Jenkins. (2007, 7th December). Man trolled the web for girls: Cops - crime - canoe.ca. [Online]. *2010(03/09/2010),* Available: http://cnews.canoe.ca/CNEWS/Crime/2007/12/07/4712680-sun.html.

[14] J. Leyden. (2010, 3rd March). How FBI, police busted massive botnet • the register. [Online]. *2010(03/09/2010),* pp. 1. Available: http://www.theregister.co.uk/2010/03/03/mariposa_botnet_bust_analysis/.

[15] Anonymous (1999, 12th February). BBC news | UK | glitter jailed over child porn. [Online]. *2010(03/09/2010),* pp. 1. Available: http://news.bbc.co.uk/1/hi/uk/517604.stm.

[16] D. Goodin. (2010, 16 August). Man sentenced for DIY gift-card cloning • the register. [Online]. *2010(16/08/2010),* pp. 1. Available: http://www.theregister.co.uk/2010/08/14/gift_card_cloning_sentence/.

7.4 Child Protection

[1] J. Moore. (2008, 27 Feb). Jane moore | make these vile monsters suffer | the sun |News|Columnists|Jane moore. [Online]. *2009(28/05/2009),* pp. 1. Available: http://www.thesun.co.uk/sol/homepage/news/columnists/jane_moore/875869/Jane-Moore-Make-these-vile-monsters-suffer.html

[2] J. Wolak, D. Finkelhor, K. Mitchell and M. Ybarra, "Online "Predators" and Their Victims Myths, Realities and Implications for Prevention and Treatment," *American Psychologist,* vol. 63, pp. 111-128, March 2008. 2008.

[3] F. Furedi. (2001, 25 April 2001). Frank furedi: Making sense of parental paranoia. [Online]. *2008(02/07/2008),* Available: http://www.frankfuredi.com/index.php/site/article/112/

[4] World Health Organization (1993), International Statistical Classification of Diseases and Related Health Problems: ICD-10 Section F65.4: Paedophilia

[5] American Psychiatric Association. (2003, Medical library | medem.com. [Online]. 2009(06/06/2009), Available: http://www.medem.com/medlib/article/ZZZU.

[6] P. Tromovitch, "Manufacturing Mental Disorder by Pathologizing Erotic Age Orientation: A Comment on Blanchard et al. (2008) [Letter to the editor]," *Archives of Sexual Behavior,* vol. 38, 2009.

[7] Anonymous BBC NEWS | technology | online child abuse images warning. [Online]. *2009(23/02/2009),* Available: http://news.bbc.co.uk/1/hi/technology/7904607.stm

[8] J. Ozimek. (2008, 9th October). Porn, abuse, depravity - and how they plan to stop it • the register. [Online]. *2009(25/02/2009),* Available: http://www.theregister.co.uk/2008/10/09/policing_internet_one/

[9] R. Clayton. (2005, November). Anonymity and traceability in cyberspace (chapter 7 - the BT 'CleanFeed' system and the failings of traceability). [Online]. *2010(12/15/2010),* pp. 1-189. Available: http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-653.pdf.

[10] C. Williams. (2009, 25 February). Small ISPs reject call to filter out child abuse sites • the register. [Online]. *2009(16/06/2009),* Available: http://www.theregister.co.uk/2009/02/25/iwf_small_isps/

[11] R. Cellan-Jones. (2009, 23 February 2009). BBC - dot.life: Can we block child abuse sites? [Online]. *2009(16/06/2009),* Available: http://www.bbc.co.uk/blogs/technology/2009/02/can_we_block_child_abuse_sites.html

[12] C. Williams. (2009, 17 June). Tories research increased net censorship • the register. [Online]. *2009(17/06/2009),* Available: http://www.theregister.co.uk/2009/06/17/tories_iwf/

[13] Anonymous Coroners and justice bill: 3 feb 2009: Public bill committees (TheyWorkForYou.com). [Online]. *2009(26/03/2009),* Available: http://www.theyworkforyou.com/pbc/2008-09/Coroners_and_Justice_Bill/02-0_2009-02-03a.1.0?s=speaker:11480#g1.278

[14] K. Lanning, "Child molesters: A behavioral analysis," National Centre for Missing & Exploited Children, Tech. Rep. Fourth Edition, September 2001.

[15] C. McCarthy. (2009, 10 Feb). Whee! new numbers on social network usage | webware - CNET. [Online]. *2009(08/07/2009),* Available: http://news.cnet.com/8301-17939_109-10160850-2.html

[16] A. Smith. (2007, 14 October). Teens and online stranger contact. Pew Internet & American Life Project, Tech. Rep. [Online] http://www.pewinternet.org/Reports/2007/Teens-and-Online-Stranger-Contact.aspx

[17] T. Lordan, D. Finkelhor, M. Ybarra, A. Lenhart, d. boyd and Audience. (2007, 3 May), Internet Advisory Committee [panel discussion] pp. 1-34. [Online] http://www.ccoso.org/library%20articles/Just%20the%20Facts%20About%20Online.pdf

[18] S. Livingstone and E. Helsper, "Taking risks when communicating on the Internet: the role of offline social-psychological factors in young people's vulnerability to online risks," *Information, Communication & Society,* vol. 10, pp. 619-644, October 2007. 2007.

[19] Internet Safety Technical Taskforce. (2008, 31 December 2008). Enhancing child safety and online technologies. The Berkman Center for Internet and Society, Harvard University. [Online]. Available: http://cyber.law.harvard.edu/pubrelease/isttf/

[20] S. Jones. (2009, 4 February). MySpace: 90,000 sex offenders removed from site | technology | guardian.co.uk. [Online]. 2009(07/08/2009), Available: http://www.guardian.co.uk/technology/2009/feb/04/myspace-social-networking-sex-offenders.

[21] d. boyd. (2009, 6 February). Apophenia: Doing the math on MySpace and registered sex offenders. [Online]. *2009(11/08/2009),* pp. 1. Available: http://www.zephoria.org/thoughts/archives/2009/02/06/doing_the_math.html

[22] Human Rights Watch. (2007, 11 September). US: Sex offender laws may do more harm than good | human rights watch. [Online]. *2009(22/09/2009),* Available: http://www.hrw.org/en/news/2007/09/11/us-sex-offender-laws-may-do-more-harm-good

[23] N. Willard, "Research that is “Outdated and inadequate?” an analysis of the pennsylvania child predator unit arrests in response to attorney general criticism of the berkman task force report," Center for Safe and Responsible Internet Use, 26 January, 2009.

[23] N. Willard. (2007, 1 September). Education world ® technology center: Nancy willard: Cyber savvy: What’s not working. [Online]. 2009(13/08/2009), pp. 1. Available: http://www.education-world.com/a_tech/columnists/willard/willard004.shtml.

[24] B. Schneier. (2007, 25 January). Schneier on security: In praise of security theater. [Online]. *2009(01/05/2009),* pp. 1. Available: http://www.schneier.com/blog/archives/2007/01/in_praise_of_se.html

[25] National Research Council, "Technical, businessand legal dimensions of protecting children from pornography on the internet," in *Committee to Study Tools and Strategies for Protecting Kids from Pornography and their Applicability to Other Inappropriate Internet Content,* 2002, pp. 1-145.

[26] C. Williams. (25th March, 2009). Sex crime 'lie detector' pilot could prompt wider use • the register. [Online]. *2009(27/03/2009),* pp. 1. Available: http://www.theregister.co.uk/2009/03/25/polygraph_trial/

[27] NATIONAL RESEARCH COUNCIL OF THE NATIONAL ACADEMIES. (2003). The polygraph and lie detection. National Academies Press, Washington, DC. [Online]. Available: http://books.nap.edu/openbook.php?record_id=10420&page=212

[28] Office of Technology Assessment. (1983, November). Scientific validity of polygraph testing: A research review and evaluation. US Congress, Washington DC. [Online]. Available: http://www.fas.org/sgp/othergov/polygraph/ota/index.html

[29] S. Aftergood. (2000, Nov 3). ESSAYS ON SCIENCE AND SOCIETY: Polygraph testing and the DOE national laboratories. *Science 290(5493),* pp. 939-940.

[30] P. Langan, E. Schmitt and M. Durose, "Recidivism of sex offenders released from prison in 1994," US Department of Justice, Bureau of Justice Statistic, Washington DC, November, 2003.

[31] T. Kendall, "Pornography, rapeand the internet," Clemson University, Clemson, South Carolina, March, 2007.

[32] M. Diamond. "Pornography, public acceptance and sex related crime: A review" *International Journal of Law and Psychiatry,* vol. 32, pp. 304-314, 8 August. 2009.

[33] B. Schneier. (2003). *Beyond Fear: Thinking Sensibly about Security in an Uncertain World.* Available: http://isbndb.com/d/book/beyond_fear_a01

[34] J. Ozimek. (2008, 14 July). Criminal record checks could hit over 14 million people | the register. [Online]. *2008(20/08/2008),* pp. 5. Available: http://www.theregister.co.uk/2008/07/14/crb_checks_total_analysis/

[35] M. Bichard. (2004, 22 June). The bichard inquiry report. [Online]. *2010(10/02/2010),* pp. 1-196. Available: http://police.homeoffice.gov.uk/publications/operational-policing/bichard-inquiry-report.html

[36] J. Ozimek. (2009, 14 December). Gov retreats on vetting database but ain't climbing down • the register. [Online]. *2010(10/02/2010),* pp. 2. Available: http://www.theregister.co.uk/2009/12/14/vetting_climbdown/

[37] C. Williams. (2010, 7 January). Anti-paedo vetting boss warns against relying on databases • the register. [Online]. *2010(10/02/2010),* Available: http://www.theregister.co.uk/2010/01/07/singleton_vetting/

[38] J. Ozimek. (2008, 20 October). UK.gov says: Regulate the internet • the register. [Online]. *2010(11/02/2010),* Available: http://www.theregister.co.uk/2008/10/20/government_internet_regulation/

[39] The YouTube Team. (not dated). YouTube Community Guidelines [Online]. *2010(11/02/2010),* Available: http://www.youtube.com/t/community_guidelines

[40] A. Modine. (2010, 10 February). YouTube saves dumb children from offensive content • the register. [Online]. *2010(11/02/2010),* Available: http://www.theregister.co.uk/2010/02/10/youtube_safety_mode/

[41] M. L. Ybarra and K. J. Mitchell. (2008, Feb). How risky are social networking sites? A comparison of places online where youth sexual solicitation and harassment occurs. *Pediatrics 121(2),* pp. e350-7.

[42] M. Ybarra and K. Mitchell, "Online aggressor/target, aggressorsand targets: a comparison of associated youth characteristics," *Journal of Child Psychology and Psychiatry*, vol. 45, pp. 1308-1316, 2004.

[43] K. Mitchell, D. Finkelhor and J. Wolak, "Risk Factors for and Impact of Online Sexual Solicitation of Youth," *Journal of American Medical Association,* vol. 285, pp. 3011-3014, 20 June. 2001.

[44] J. Palfrey, J. boyd, d. Sacco and L. DeBonis, "ISTTF: Enhancing Child Safety and Online Technologies," 3 Feb. 2009.

[45] N. Willard, "Parent guide to Cyberbullying and Cyberthreats," Center for Safe and Responsible Internet Use, 2007.

[46] J. Ozimek. (2009, 19th January). Govt uses obscenity law to stuff up cartoon sex loophole • the register. [Online]. *2010(28/01/2010),* Available: http://www.theregister.co.uk/2009/01/19/evil_cartoon_badness/

[47] J. Ozimek. (2009, 17th March). Put down your pens: Cartoons next on censor block • the register. [Online]. *2009(25/03/2009),* Available: http://www.theregister.co.uk/2009/03/17/cartoon_badness/

[48] Anonymous Coroners and justice bill: 3 feb 2009: Public bill committees (TheyWorkForYou.com). [Online]. *2009(26/03/2009),* Available: http://www.theyworkforyou.com/pbc/2008-09/Coroners_and_Justice_Bill/02-0_2009-02-03a.1.0?s=speaker:11480#g1.278

[49] J. Ozimek. (18th February, 2009). UK 'bad' pics ban to stretch? • the register. [Online]. *2009(25/03/2009),* Available: http://www.theregister.co.uk/2009/02/18/cartoon_law_loophole_choke/

[50] J. Ozimek. (2009, 29 November). Cartoon smut law to make life sucky for olympic organisers • the register. [Online]. *2010(01/03/2010),* pp. 2. Available: http://www.theregister.co.uk/2009/11/29/olympic_logo_lisa_simpson/

[51] E. Qualye. (2008, September). The COPINE project. *Irish Probation Journal* [Online]. *2010(02/03/2010),* pp. 65-83. Available: http://www.probation.ie/pws/websitepublishing.nsf/Content/Irish+Probation+Journal

[52] J. Oates. (2008, 8 December). Aussie convicted over simpsons sex pics • the register. [Online]. *2010(03/03/2010),* pp. 1. Available: http://www.theregister.co.uk/2008/12/08/simpsons_supreme_court/

[53] J. Oates. (2010, 28 January). Aussie man convicted for simpsons smut • the register. [Online]. *2010(03/03/2010),* pp. 1. Available: http://www.theregister.co.uk/2010/01/28/australia_simpsons/

[54] K. McCarthy. (2001, 5 January). Royal & sun alliance sacks ten over obscene emails • the register. [Online]. *2010(03/03/2010),* pp. 1. Available: http://www.theregister.co.uk/2001/01/05/royal_sun_alliance_sacks_ten/

[55] K. McCarthy. (2000, 18 December). More email victims at royal & SunAlliance • the register. [Online]. *2010(03/03/2010),* pp. 1. Available: http://www.theregister.co.uk/2000/12/18/more_email_victims_at_royal/

[56] J. Wolak, K. J. Mitchell and D. Finkelhor. (2007, Dec). Does online harassment constitute bullying? an exploration of online harassment by known peers and online-only contacts. *J. Adolesc. Health 41(6 Suppl 1),* pp. S51-8.

[57] E. Barnett. (2009, 19 November). Can bebo's panic button beat cyber-bullying? - telegraph. [Online]. *2010(07/04/2010),* Available: http://www.telegraph.co.uk/technology/social-media/6600032/Can-Bebos-panic-button-beat-cyber-bullying.html

[58] C. Williams. (2010, 18 March). Facebook faces home sec over lack of 'panic button' • the register. [Online]. *2010(07/04/2010),* Available: http://www.theregister.co.uk/2010/03/18/facebook_johnson/

[59] R. Cellan-Jones. (2009, 18 November). BBC - dot.life: Facebook v ceop. [Online]. *2010(07/04/2010),* Available: http://www.bbc.co.uk/blogs/technology/2009/11/facebook_v_ceop.html

[60] Anonymous (1959, 29th July). Obscene publications act 1959 (c.66) - statute law database. [Online]. *2010(26/05/2010),* Available: http://www.statutelaw.gov.uk/content.aspx?activeTextDocId=1128038.

[61] Anonymous Bbfc - the official website of the bbfc. classification for entertainment, movies and video games. [Online]. *2010(26/05/2010),* Available: http://www.bbfc.co.uk/about/.

[62] J. Ozimek. (2009, 30th June). UK obscenity law: Where to now? • the register. [Online]. *2010(27/05/2010),* pp. 2. Available: http://www.theregister.co.uk/2009/06/30/obscenity_law_where_now/.

[63] Anonymous Criminal justice and immigration act 2008 (c. 4). [Online]. *2010(27/05/2010),* Available: http://www.opsi.gov.uk/acts/acts2008/ukpga_20080004_en_9#pt5-pb1-l1g63.

[64] J. Wolak, K. Mitchell and D. Finkelhor, "THE EXPOSURE OF YOUTH TO UNWANTED SEXUAL MATERIAL ON THE INTERNET A Nation Survey of Risk, Impact and Prevention." *Youth & Society*, vol. 34, pp. 330-358, March. 2003.

[65] M. Flood and C. Hamilton, "Youth and pornography in australia: Evidence on the extent of exposure and likely effects," The Australia Institute, Tech. Rep. 52, February 2003.

[66] T. Byron. (2008, 27 March). Department for children, schools and families : Byron review. Dept Children, Schools and Families. [Online]. Available: http://www.dcsf.gov.uk/byronreview/index.shtml.

[67] M. Seamark. (2010, 28th May). Facebook grooming: How pervert postman used site to groom hundreds of children | mail online. [Online]. *2010(28/05/2010),* Available: http://www.dailymail.co.uk/news/article-1282157/Facebook-grooming-How-pervert-postman-used-site-groom-hundreds-children.html.

[68] J. Wolak, D. Finkelhor and K. Mitchell, "1 in 7 youth: The statistics about online sexual solicitations," Crimes against Children Research Center, University of New Hampshire, December. 2007.

[69] J. Patchin and S. Hinduja, "Bullies Move Beyond the Schoolyard - A Preliminary Look at Cyberbullying," *Youth Violence and Juvenile Justice,* vol. 4, pp. 148-169, April, 2006.

[70] Anonymous, Crime: Protecting pennsylvania against crime - pennsylvania office of attorney general. *2010(01/06/2010)*

[71] K. Mitchell, J. Wolak and D. Finkelhor, "Trends in Youth Reports of Sexual Solicitations, Harassment and Unwanted Exposure to Pornography on the Internet," *Journal of Adolescent Health,* vol. 40, pp. 116-126, 26 May. 2007.

[72] Crimes Against Children Research Center. First youth internet safety survey (YISS-1). [Online]. *2009(30/10/2009),* pp. 1. Available: http://www.unh.edu/ccrc/youth_internet_safety_survey.html.

[73] Crimes Against Children Research Center. Second youth internet safety survey (YISS-2). *2010(21/07/2010)* .

[74]] L. Campher and C. Bezuidenhout, "A criminological overview of paedophile activities on the internet," South African Professional Society on the Abuse of Children, University of Pretoria, Tech. Rep. 1, 2007, 2007.

[75] M. Campbell. (2005, Cyber bullying: An old problem in a new guise? *Australian Journal of Guidance and Counselling* [Online]. *15(1),* pp. 68-76. Available: http://www.atypon-link.com/AAP/doi/abs/10.1375/ajgc.15.1.68.

[76] ASACP: Association of Sites Advocating Child Protection. Industry best practices. [Online]. *2010(22/07/2010),* pp. 1. Available: http://www.asacp.org/index.php?content=best_practices&PHPSESSID=a41065591cb64fea95c37d9cfdab0757.

[77] T. Lordan, D. Finkelhor, M. Ybarra, A. Lenhart, d. boyd and Audience, "Internet Caucus Advisory Committee [panel discussion]," pp. 1-34, 3rd May. 2007.

[78] K. Subrahmanyam, P. M. Greenfield and B. Tynes, "Constructing sexuality and identity in online teen chat room," *Applied Developmental Psychology*, vol. 25, pp. 651-666, 2004.

[79] J. Carr. "Child abuse, child pornography and the internet," NCH, London, Tech. Rep. 1, December 2003.

[80] D. Lazarová. (9th Feb, 2007). Child porn consumers safe from prosecution in the czech republic - radio prague. [Online]. *2010(29/07/2010),* pp. 1. Available: http://www.radio.cz/en/article/88189.

[81] CEOP. Thinkuknow - 11-16s - what can go wrong? - chat rooms and IM. *2010(29/07/2010).*

[82] CEOP. Thinkuknow - 11-16s - what can go wrong? - social networking sites. *2010(29/07/2010).*

[83] J. Wolak, K. Mitchell and D. Finkelhor, "Escaping of connecting? Characteristics of youth who form close online relationship," *Journals of Adolescence,* vol. 26, pp. 105-119, 2003.

[84] C. Williams. (13th April, 2010). Facebook rejects CEOP 'panic button' demands (again) • the register. [Online]. *2010(30/07/2010),* pp. 1. Available: http://www.theregister.co.uk/2010/04/13/ceop_facebook_dc/.

[85] R. Trenholm. (12 July, 2010). Facebook ClickCEOP child-protection app: Just don't call it a panic button | crave | CNET UK. [Online]. *2010(30/07/2010),* pp. 1.

Available: http://crave.cnet.co.uk/software/facebook-clickceop-child-protection-app-just-dont-call-it-a-panic-button-49306151/.

[86] Slack and P. Sims. (12 March, 2010). Facebook warning after peter chapman admits ashleigh hall murder | mail online. [Online]. *2010(30/07/2010),* pp. 1. Available: http://www.dailymail.co.uk/news/article-1256307/Facebook-warning-Peter-Chapman-admits-Ashleigh-Hall-murder.html.

[87] N. Willard and M. Novotny. (2002, *Computer Ethics, Etiquette and Safety for the 21St-Century Student* Available: http://isbndb.com/d/book/computer_ethics_etiquette_and_safety_for_the_21st_century_st.

[88] J. Oates. (2009, 19 March). Secret aussie blacklist leaked • the register. [Online]. *2010(03/08/2010),* pp. 1. Available: http://www.theregister.co.uk/2009/03/19/australia_list_leaked/.

[89] Anonymous "Joint Statement on Internet Censorship," pp. 1, July, 2009. [Online] Available: http://www.getup.org.au//files/campaigns/jointstatementinternetcensorship.pdf

[90] T. Martin, C. Durbin, M. Pawlewski and D. Parish. A Pseudonymous Peer-2-Peer Review System for Child Protection On-line *Journal of International Commercial Law and Technology* [Online]. *5(2),* pp. 04 August 2010. Available: http://www.jiclt.com/index.php/jiclt/article/viewArticle/104.

[91] Crisp Thinking, "CRISP THINKING'S ANTI-GROOMING TECHNOLOGY ACHIEVES 98.4% ACCURACY IN INDEPENDENT TESTING [Press release]," 14 January. 2008.

[92] M2 Presswire. (2008, 14 Jan 2008). Crisp thinking's anti-grooming technology achieves 98.4% accuracy in independent testing. | M2 presswire (january, 2008). [Online]. *2008(02/04/2008),* pp. 1. Available: http://www.accessmylibrary.com/coms2/summary_0286-33698195_ITM

[93] Anonymous (2010, 13 May). IWF reveals commercial core of child sexual abuse ‘brands’ on the internet. [Online]. *2010(05/08/2010),* pp. 1. Available: http://www.iwf.org.uk/media/news.285.htm.

[94] L. Falshaw, C. Friendship and A. Bates, "Assessing reconviction, re-offending and recidivism in a sample of sexual offenders." *Home Office Research, Development and Statistics Directorate Research Findings no. 183*, 2003.

[95] J. Oates. (2010, 13 April). Thousands wrongly labelled by CRB checks • the register. [Online]. *2010(09/08/2010),* pp. 1. Available: http://www.theregister.co.uk/2010/04/13/crb_checks_wrong/.

[96] J. Ozimek. (2008, 17 August). Malicious gossip could cost you your job • the register. [Online]. *2010(09/08/2010),* pp. 1. Available: http://www.theregister.co.uk/2008/08/17/gossip_work_check/.

[97] J. Ozimek. (2009, 11 December). Headteachers slam 'disproportionate' vetting database • the register. [Online]. *2010(09/08/2010),* pp. 1. Available: http://www.theregister.co.uk/2009/12/11/headteachers_slam_isa/.

[98] C. Williams. (2010, 7 January). Anti-paedo vetting boss warns against relying on databases • the register. [Online]. *2010(10/02/2010),* Available: http://www.theregister.co.uk/2010/01/07/singleton_vetting/.

[99] A. Flood. (2009, 10 July). Authors in revolt against plans to vet them for school visits | books | guardian.co.uk. [Online]. *2010(10/08/2010),* pp. 1. Available: http://www.guardian.co.uk/books/2009/jul/10/authors-vet-school-visits.

[100] F. Furedi and J. Bristow. (2008, *Licensed to Hug: How Child Protection Policies are Poisoning the Relationship between the Generations* Available: http://isbndb.com/d/book/licensed_to_hug.

[101] L. Magid. (2010, 12 January). Is taser's phone-monitoring product overparenting? | CES 2010 - CNET. [Online]. *2010(11/08/2010),* pp. 1. Available: http://ces.cnet.com/8301-31045_1-10433539-269.html.

[102] Anonymous Lessons from the dot-com bubble. [Online]. *2010(13/08/2010),* pp. 1. Available: http://www.theinvestorsjournal.com/lessons-from-the-dot-com-bubble/.

[103] J. Ozimek. (2008, 14 July). Criminal record checks could hit over 14 million people | the register. [Online]. *2008(20/08/2008),* pp. 5. Available: http://www.theregister.co.uk/2008/07/14/crb_checks_total_analysis/

[104] Anonymous (1998, Self-reported honesty among middle and high school students responding to a sexual behavior questionnaire. *Journal of Adolescent Health 23(1),* pp. 20.

[105] Young, K. (2005). Profiling online sex offenders, cyber-predators and pedophiles. *Journal of Behavioural Profiling, 5*(1)

[106] M. Newman. (2009, 10 September). Times higher education - paedophilia research riles and titillates the academy. [Online]. *2009(12/10/2009),* pp. 1. Available: http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=408084&c=2

[107] J. Oates. (2009, 22 September). Home office stonewalls ID findings • the register. *2009(15/10/2009)*

[108] J. Oates. (2009, 30 September). Home office declines to detail DNA-for-foreigns trial • the register. *2009(15/10/2009)*

[109] B. Glassner. (1999, *The Culture of Fear: Why Americans are Afraid of the Wrong Things* Available: http://isbndb.com/d/book/the_culture_of_fear_a01.

[110] S. Moeller. (1999, August). *Compassion Fatigue : How the Media Sell Disease, Famine, War and Death* (1st ed.) .

[111] Kable. (2010, 28 April). NPfIT ignored NHS culture, says halligan • the register. *2010(9/28/2010),* Available: http://www.theregister.co.uk/2010/04/28/npfit_halligan/.

[112] [Versign]. (2010, 23 December). The total number of web domain names. [Online]. 2011(3/15/2011), Available: http://www.labnol.org/internet/total-web-domain-names/18395/.

[113] Anonymous (2008, 25 July). Official google blog: We knew the web was big... [Online]. 2011(3/15/2011), Available: http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html.

8. APPENDIX A – TERMINOLOGY

Bayesian refers to a statistical method to determine probability.

Chat a blanket term that describes real-time Internet communication, this includes IM, IRC, SNS chat functions etc.

Gaussian Mixture Model a clustering technique used in statistical probability.

Emoticons aka smileys graphical representations of emotions e.g. happy, sad, confused, embarrassed.

Microsoft Network (MSN) a blanket term that may refer to the chat service, the protocol (underlying the chat service) or the application(s) i.e. the user interface to the chat service.

Identity Provider (IdP) an authority where users register their details, the IdP then provides credentials and attests to the user's identity to relying parties.

Instant Messaging (IM) a form of two-way communication between two or more individuals conducted in real-time. Traditionally text based, many instant messaging platforms provide support for audio and video.

Internet Relay Chat (IRC) a form of Internet chat based around channels, similar to chat room is supports group chat or one-to-one communication.

Relying Party (RP) a website, service, platform or application that accepts a users identity based on assurances from a IdP rather than collecting and managing user details itself.

Social Network Site (SNS) a platform that allows users to publish content in the form of a profile, typically mood/status, images, video and links. They user can choose to

make the profile public, visible only to approved users (friends) or somewhere in between.

Token a unit of data created from features extracted from text. Tokens can be used to train a model or classified (tested against a model).

User Centric an ethos or approach that places the user in control of how their personal data is used.

9. APPENDIX B – RISKS AND SAFETY ON THE INTERNET

In the main body of the thesis the author calls for a broad survey of UK children similar to the US YISS project.

Since completing this work the author has become aware of the EU Kids Online project a trans -European project that surveyed 23,420 children across 25 countries [1].

The report broadly reflects the findings of this thesis, with one key difference – the problem of cyberbullying. EU Kids Online found that 19% of children had engaged in bullying in the past 12 months, 5% report being bullied on a weekly basis. Of the 5% who have been bullied online 57% were very upset or fairly upset, however 94% got over the bullying immediately or within a few days.

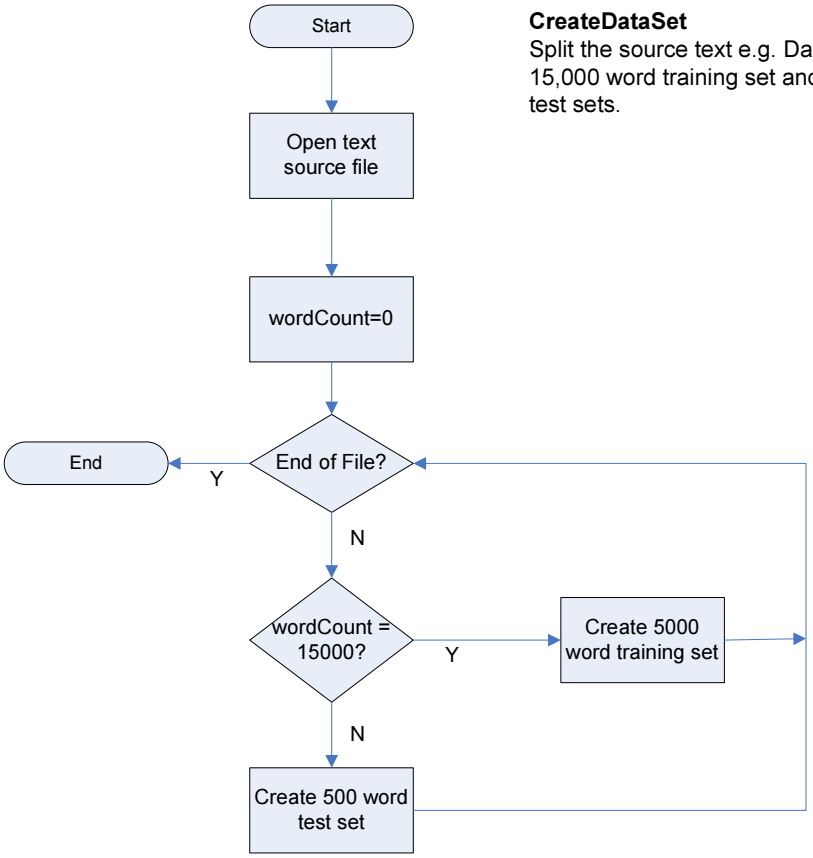
Given the currency, sample size and methodology of the EU Kids Online project the author recognises this is probably more representative of state of cyberbullying in the UK.

[1] S. Livingstone, L. Haddon, A. Gorzig and K. Olafsson. (2010, 21 October). Risks and safety on the internet - EU kids online. EU / LSE. [Online]. Available: http://www2.lse.ac.uk/media@lse/research/EUKidsOnline/EUKidsII%20%282009-11%29/Survey/Survey%20documents.aspx

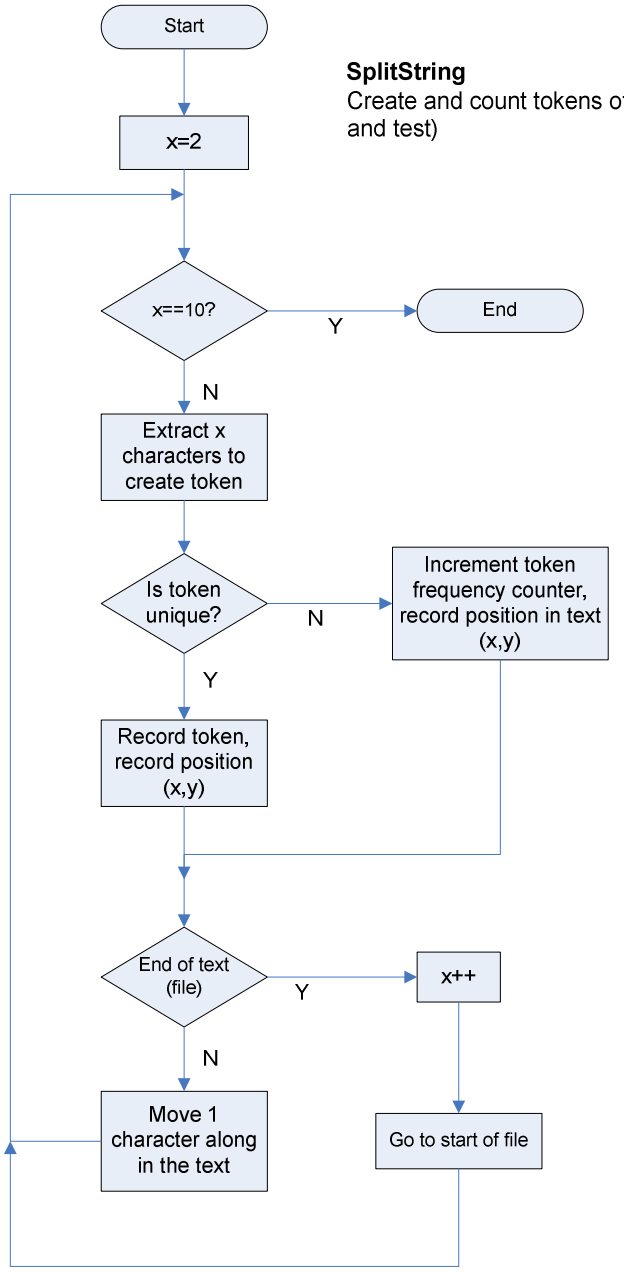
10. APPENDIX C – FLOWCHARTS

The following flowcharts were created as part of the software engineering process and describe how the applications work.

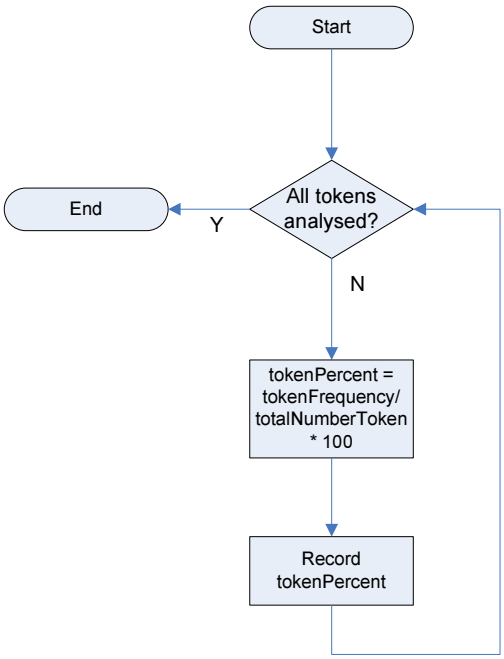
10.1 Text Analysis



CreateDataSet
Split the source text e.g. Darwin into a single 15,000 word training set and multiple 500 word test sets.

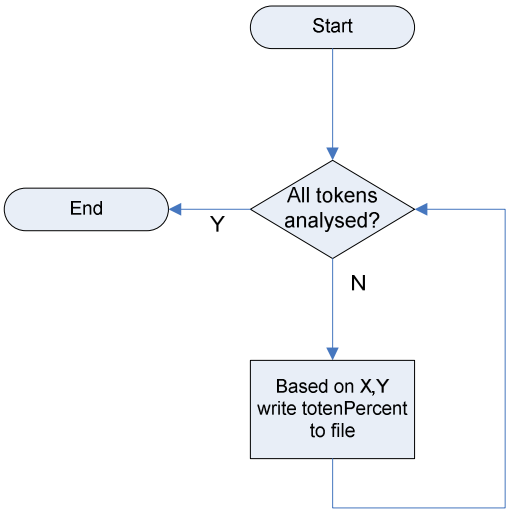


SplitString
Create and count tokens of each set (training and test)

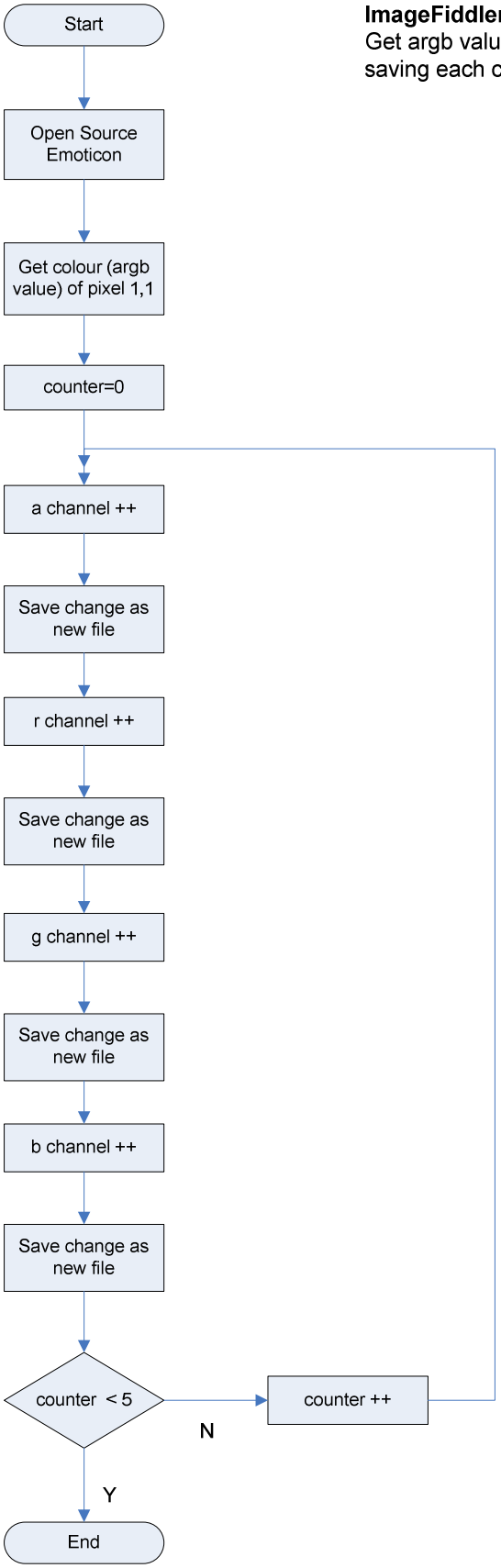


Percent
Determine the frequency, as a percentage of each token

LoadUp
Each vector consists of the percentage of tokens 2 to 10 characters in length.
Based on the X,Y (X = number of characters, Y = position of token in the text e.g. 10 characters from start of file/set) write tokenPercent, to create token vectors.

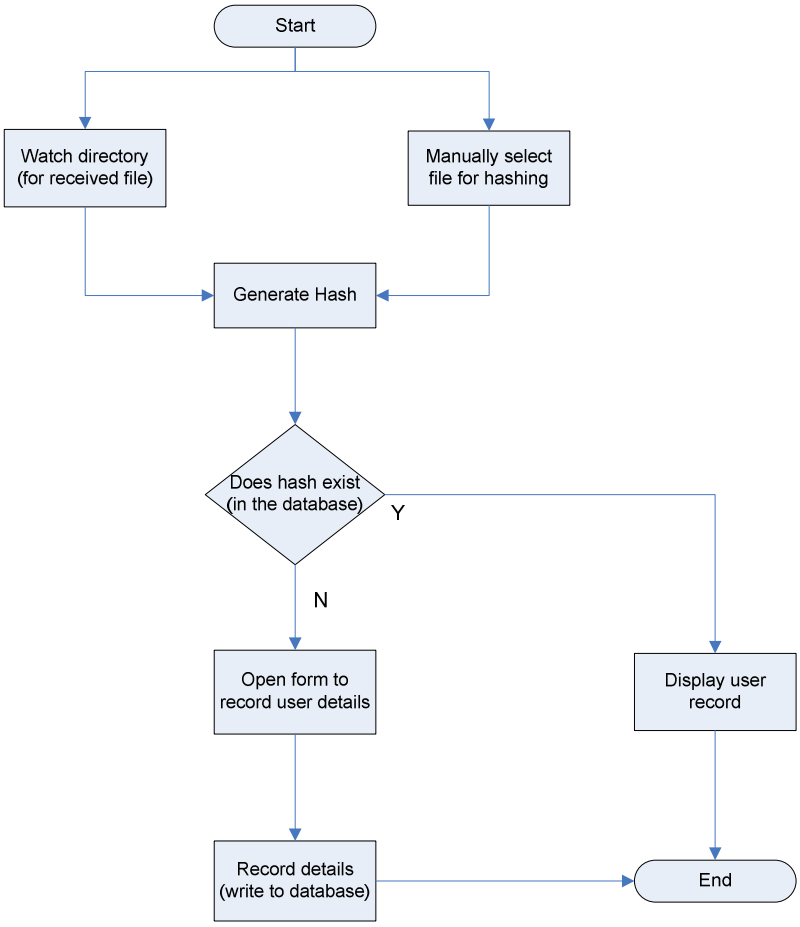


10.2 Steganographed Custom Emotions



ImageFiddler (create the SCE set)
Get argb value, and alter +/- 5 in each channel, saving each change as a new file (emoticon)

HashChecker
Generate or validate hash of an emoticon.
Record or display user details.



11. APPENDIX D - CUSTOM EMOTICON CHANGE – EXPERIMENT LOG

Original Source Emoticon: Baring_teeth_smiley.png

Manufactured

File name	Hash	Colour
1.png	abbd77ad7b06513172c1c58e8086e5e4	[A=1,R=1,G=1,B=1]
2.png	e891351cf88b0f115ef5c66ce7a9fde4	[A=2,R=1,G=1,B=1]
3.png	00fb459af4ef96ed36acc10f54346643	[A=3,R=1,G=1,B=1]
4.png	46bb1f29fc247fc090a0fc23d5c8ea05	[A=4,R=1,G=1,B=1]
5.png	c401d91c6de43a2a254f8223db412468	[A=5,R=1,G=1,B=1]
6.png	2ff95002bb5b1aedda0cc71d2950fff4	[A=6,R=1,G=1,B=1]
7.png	2c1e17ec897242e708c3379714e28c96	[A=7,R=1,G=1,B=1]
8.png	cb4fe8a3630eb8416736a8ec399567a7	[A=8,R=1,G=1,B=1]
9.png	ba60adff743e9cf956e38a2b0f85e618	[A=9,R=1,G=1,B=1]
10.png	212570f504760ebcbd49400a07659741	[A=10,R=1,G=1,B=1]
11.png	de4739e58b347613346352ad44b46bbc	[A=11,R=1,G=1,B=1]
12.png	d679442dee61de6d6a979bd011e233fd	[A=12,R=1,G=1,B=1]
13.png	81403885e9191cc41c7ea6970e7085fd	[A=13,R=1,G=1,B=1]
14.png	743c50c297d980ce0cccf19889148e14	[A=14,R=1,G=1,B=1]
15.png	cd19dc374bb4b5a66ff89865da20f50d	[A=15,R=1,G=1,B=1]
16.png	b962c04ddeb74ba98562ef84b383dcfc	[A=16,R=1,G=1,B=1]
17.png	85f535dc8cdf3f825deda497490bf013	[A=17,R=1,G=1,B=1]
18.png	5babc08c45fcb6215f8bda2e5f426ff4	[A=18,R=1,G=1,B=1]
19.png	d5207f586e3c6758657e26894c0bd030	[A=19,R=1,G=1,B=1]
20.png	96a86399862d3842de198b89405d2b0f	[A=20,R=1,G=1,B=1]
250.png	55a303df5253f22fc6f6ad94f71b4d10	[A=255,R=2,G=5,B=5]
251.png	4270c7716bdddd23c3703e3e4b69a45b	[A=255,R=3,G=1,B=1]

Sent from source - Received at first hop

Filename	Hash	Colour
XZ9Ot0TmxdtD8FVKdCzBjlqZ5MM=	abca826720df868d736f9e805af5a0ea	[A=1,R=0,G=0,B=0]
S0haOruopUs61SxoHb9e0SypyOA=	513540dac8184ab6e9adbb0c12efedcb	[A=2,R=0,G=0,B=0]
iW26K6M5v3siMBTQs+TGfx3TCSg=	138e4f0c180827e123f4c810524d9eb2	[A=3,R=0,G=0,B=0]
wa+6L+YRj9+cyZYMu6r2FCo+3bvI=	8923b51d404481ec7699a3ad71531767	[A=4,R=0,G=0,B=0]
YC82Fr0uMPPkNwxLzarQE7ul1BYI=	ca321ec838da004eab31f2381ff59ee5	[A=5,R=0,G=0,B=0]
+152b1+uAQqZpcB5twyp69wFKKc=	bde77fbbd7fce0ac6226945e78d643ef	[A=6,R=0,G=0,B=0]
fi3eBVjJVr04yKOUFuTo6tUUDlQ=	82214c7050defd6ecbc8d8b52e6f175d	[A=7,R=0,G=0,B=0]
n+jK+kCmD+jVkAgT0mq0qbPccNE=	ba0e1ce87bbc5ac43a199ba9cac68eaf	[A=8,R=0,G=0,B=0]
0uOqkPfD7Qk8gjmgyph+gOhUNTk=	d308765676fee3b7b0251857c5205b2d	[A=9,R=0,G=0,B=0]
JrYMt2IRJI2YfJ8w59MDW2FEXb+Y=	31e8bce24d9429599a33c29fd3e09db6	[A=10,R=0,G=0,B=0]
zu33HVQYRHOJysJKEADwfc8wQ9Q=	47460cf43c8727ceb568820e6f8b1f40	[A=11,R=0,G=0,B=0]
73lDU0LrwA6HBSqR472RSMh0vD4=	fb35e42fb357a94a4e6f6aa337d896a3	[A=12,R=0,G=0,B=0]
2QLZz2n1GYUNebyEpqNgAkLTAeo=	e26ca8b51d3ad485baaf0e9cecce42d7	[A=13,R=0,G=0,B=0]
3RCgylSpqWP4dRniziBXYfN5l8A=	f785c8dcd09acecbc5734f002914987d	[A=14,R=0,G=0,B=0]
FYUYQEeeB2F2VUQkWx2FZWgY2FpZps=	ad25d8b8141ed3478304190b5c9d230b	[A=15,R=0,G=0,B=0]
Ea9SlCOVLabBovoXrvp+aFQKdoM=	38166709276e4be3b20632244e552a53	[A=16,R=0,G=0,B=0]
YUjacEZEkgXho4Glkk2Fz40UgWk=	7f960362080863065825476b17898908	[A=17,R=0,G=0,B=0]
jJFtLvblSFwOTVja5TJqS5vHJ5I=	95b3710ff12fcc4f7bb64dc9f6f7ca73	[A=18,R=0,G=0,B=0]
LERqMKWlIJR2Y9LpKzXaVvG42FeA=	a0c872cfa7dcf7b25d43b2c4db80c046	[A=19,R=0,G=0,B=0]
O1G2FsKjA9IpjU0w407ouiy+fck0=	ad9e52fdfea193b30e22316589ef5450	[A=20,R=0,G=0,B=0]
xoSBwJd5hg6o2FU66mQmqb64w58k=	a68f44f1f022123bdc5225f96ee7a44c	[A=255,R=2,G=5,B=5]

20fyZYlfgY9JTxYVYUHhAQCrVOc=	0a87129822022c25a3f95525b07a4fd0	[A=255,R=3,G=1,B=1]

Recovered at first hop - Sent to 3rd Party (received at 3rd Party)

File name	Hash	Colour
8+srWWpU1k9kienyGg9qL2F1RWqY=	8dbee777d14ff11e224090ff51d0bba9	[A=1,R=0,G=0,B=0]
AHDhk87Dg7A2FFeZhZAhFzNzUj3w=	768307fa221a1942d977198f4ec205de	[A=2,R=0,G=0,B=0]
NLn5sPdrphC2GP427HLP0fyeS6I=	05fcd23456b5e78f3c788979b37104bc	[A=3,R=0,G=0,B=0]
36wrw7N35WJnQOsLZXomYDVhIUU=	5277a9690127b53dc69555344b2bc798	[A=4,R=0,G=0,B=0]
y8KcHIF4oZBe1E2Fw8dfGRQ+PZm8=	7fe754185e61654f868583630b69988a	[A=5,R=0,G=0,B=0]
JVobDgMA4H44hZex+vEDTyZXmcc=	3461579fed83497abce017195f2cd292	[A=6,R=0,G=0,B=0]
UCdz+w2kqctb7ecHSGnPDaDP4Ts=	4b9d37dcd0ce5ea83162b126712da6c8	[A=7,R=0,G=0,B=0]
2fv7IbcHig1z4PA386TbTMzoWqg=	4d10044ae49a5de7a6260f7189502014	[A=8,R=0,G=0,B=0]
OAea4rJBht7gk0oVCFxjW7rJ2FyE=	ce19061009c2eb944c045e369aad3e67	[A=9,R=0,G=0,B=0]
LKH0LMFg816WU9GN4DXWmOdsoOE=	e31f6df201a10c4c7390e1f16bce51fc	[A=10,R=0,G=0,B=0]
FgWPdHCOZfgcXQlJX5mctdlKNpM=	6aa8596ed9e90404a50dd8996d600c44	[A=11,R=0,G=0,B=0]
Dr7+f7inzbz90bvLhVPK33fKX9U=	dabf089242c8b2e1bf25d3b7de5b68e6	[A=12,R=0,G=0,B=0]
v0ZTEjnFj79lnN2FZ1QJAMXKE0kk=	4e453bafa29e259f3f8e1ec10ded6c6e	[A=13,R=0,G=0,B=0]
Fh8EBabzoi5J6sHdzlO1jUzFig0=	c5f9592a287217fd2da321b34372c71c	[A=14,R=0,G=0,B=0]
miWUOqxl42jXek0j0j8VIc7A3vg=	35a696257622f1acc119d02a3d510e4d	[A=15,R=0,G=0,B=0]
gK4Iev2GfFGZR2X9KtEAND2Fo8I4=	1dd0e522ded24235ec10b236a956d457	[A=16,R=0,G=0,B=0]
P785in2FyWZL99l2k+4xa5mZremY=	8c854b8cc19f98780545f04bad4eb758	[A=17,R=0,G=0,B=0]
MHeJxhbH0B9w7KMq36VB+EKeVow=	53a32b7fd099cc9d6b8833fcf38abc80	[A=18,R=0,G=0,B=0]
W64bKxFjV8CkeY7TB3SUEe35Cr8=	33270e99599b6179137e1748e415752a	[A=19,R=0,G=0,B=0]
c9tFBRx7oP+ZtKhMc+zVgoFuqxo=	720174dcb239f1bea8aa1827e7244e17	[A=20,R=0,G=0,B=0]
3IO8+WnbddXcpb0LQSyyx2F5GsIU=	7d9f82ddb7a2cffededebd4948642e0f	[A=255,R=2,G=5,B=5]
fQnVoyI6PYcw8gL9DCfO2FUO4660=	97f8f4c9bdf68944716e17bc859e7ca0	[A=255,R=3,G=1,B=1]