Towards the Resolution of Safety and Security Conflicts

Catherine Menon, Stilianos Vidalis
School of Engineering and Computer Science, University of Hertfordshire, Hatfield, AL10 9AB, UK
c.menon@herts.ac.uk and s.vidalis@herts.ac.uk

***Abstract*—**Safety engineering and cyber security have complementary aims, but typically realise these using different techniques, risk assessment methods and cultural approaches. As a result, the integration of safety and cyber security concerns is a complex process, with potential for conflict. We present a generalized taxonomy of common conflict areas between safety and cyber security, oriented around the development and deployment lifecycle, and supplement this with a discussion of concepts and methodologies for resolution based on the shared principle of defence-in-depth.

Keywords*—Safety-critical systems, security*

I. Introduction

Safety and cyber security share a common aim of increasing the dependability and integrity of systems. However, they have distinctly different focuses, with safety being concerned with protection from accidental failures resulting in harm, while security is concerned with protecting a system from malicious attack [12]. Because of this difference in focus, the safety and security domains employ distinct threat assessment techniques and risk acceptance criteria.

Integrating safety and security considerations is therefore a complex task. There are a number of cultural and technical practices embedded within in the safety domain which do not translate well to cyber-security and vice versa (e.g. the primacy of the principles of diversity and redundancy). Moreover, the requirements of safety engineering and cyber security can in some cases be perceived to conflict, such as the requirement to continually update or patch a system.

Cultural perspectives have also played a role in the difficulty in adequately assuring and separating safety and security concerns. In many regards, safety and security have historically been considered to overlap, with little distinction being made between them and a relatively poor understanding of the perceived conflicts. In several natural languages the distinction is itself unclear (e.g. the Dutch *veiligheid* vs *beveiliging* [13]). As a result, it is not uncommon for developers to be unaware of the areas in which safety and security can present competing requirements, nor familiar with techniques to accommodate both disciplines.

Although existing guidance identifies some common principles of safety and security [2], the relative novelty and immaturity of this area means that the guidance is of necessity presented at a relatively high level. Consequently, at development level, it can be difficult to identify likely areas of conflict between safety and security requirements or assess the scope and impact of decisions to prioritise one over the other.

In this paper we seek to address this gap by providing a generalized development-level taxonomy of areas of conflict between safety and security, as well as identifying cultural concepts which are shared between these domains. This provides a foundation for methodologies drawing on these shared concepts to be used to minimise areas of conflict.

In Section 2 we describe some of the commonalities and differences between approaches to safety and cybersecurity and provide a foundation for Section 3, the generalized taxonomy of conflict areas. Section 4 discusses shared cultural concepts and methodologies and Section 5 concludes.

II. Safety and Security Approaches

Historically it has been common for safety engineers to assume that their systems are free from malicious interference due to isolation ("air gap") or to the use of bespoke software, specific to certain industries [8]. In recent years, however, a number of high-profile cyber security breaches on critical national infrastructure have emphasised that safety-critical systems are, by their very nature, attractive targets for malware and ransomware [6] [7]. Moreover, systems shown to be unsafe in certain ways (e.g. undefined behaviour) are also more likely to be insecure, as demonstrated in the null pointer dereference exploit to which mobile phones are vulnerable [14].

More recently, there has been a growing understanding of the differences and commonalities between safety and security requirements, particularly that a system vulnerable to security breaches is also vulnerable to safety failures as a result of these breaches ("if it's not secure, it's not safe" [5]). For example, a successful security breach may compromise effective separation between safety-critical and non safety-critical code, or may disable a secondary system used for functional redundancy.

Nevertheless, within the UK the two disciplines of safety and cyber security operate within very different regulatory and risk assessment environments. Safety is governed by the ALARP principle: the safety risk presented by a system must be reduced As Low As Reasonably Practicable [3]. This provides flexibility in the approach to reducing risk, while allowing for safety arguments based on both good practice and first principles. The focus of much safety engineering is on hazard analysis, mitigation and computation of the residual risk from either a qualitative or quantitative perspective. Using this approach, it is then possible to address hazards caused by cyber security factors in the same way as "native" safety hazards [17].

Cyber security, by contrast, does not make use of the ALARP principle but rather requires that operators of essential services take measures which are "appropriate and proportionate" [17]. Within the constraints of the applicable legislation, cyber security risk assessment and mitigation is governed by the risk appetite of the relevant organization. A significant focus of security legislation is on forensics and detection, in contrast to the emphasis in safety legislation on prevention of accidents. This takes into account the different nature of safety accidents and cyber security breaches: it is possible for a security breach to be ongoing without detection, while accidents are typically immediately visible.

Assuring the safety of a system relies on having a full and complete knowledge of the components in the system, their provenance, and their interactions. Safety cases should be developed in parallel with the system [10], and updated whenever the system changes. Modular safety cases [11] address this by providing a method of assuring individual components, although the validity of this assurance is of necessity highly restricted and dependent on non-interference between components. Safety promotes a defence-in-depth philosophy [15], encouraging the use of overlapping safety provisions so that a single failure is ideally compensated for by the succeeding layers of safety provision.

Cyber security similarly requires a full knowledge of the system components, with a further and crucial need to understand and integrate any changes or updates to these components. Like safety, cyber security also promotes a defence-in-depth approach to protect confidentiality, availability and integrity of system data and infrastructure [16]. In a constantly-changing risk environment, using the defence-in-depth principle has the added advantage of impeding or retarding any attack, so that the damage done before it has been noticed is minimized.

Although there is no fundamental conflict between the regulatory and best practice approaches used in cyber security and safety, there is also currently no expectation that they necessarily automatically work in compliance with each other. In Section 4 we will draw on the common principle of defence-in-depth when identifying potential guidance and mitigation strategies that can be used across both domains.

III. Taxonomy of Conflict Areas

In this section we identify a taxonomy of conflict areas in which the techniques, assessment methods and culture of safety and cyber security might be expected to conflict. The purpose of such a taxonomy is to aid developers and engineers in predicting where further analysis might be necessary, or to alert them to potential conflicts which may otherwise have gone unidentified. We note that inclusion of a conflict area does not imply that there will necessarily be a conflict between safety and security concerns in this area for any specific system: rather, this can be taken to imply that the general goals and techniques of safety and security in this area are liable to conflict.

This taxonomy has been designed to be generally applicable to all safety-critical systems. Individual systems must still be assessed using information specific to their properties and characteristics, and the taxonomy should not be used as a substitute for such considered assessment. The taxonomy is also applicable at development level, and hence should be used in conjunction with existing high-level guidance such as [2].

The taxonomy as presented in Table 1. It is structured around the design and deployment lifecycle, with a specific additional focus on risk approaches and the regulatory environment.

Table 1: Safety and security conflict areas taxonomy

Phase	Conflict area	Safety perspective	Security perspective
Requirements	Diversity and redundancy	Using multiple diverse redundant systems is a fundamental tenet of safety engineering used to eliminate single points of failure and provide defence-in-depth.	The presence of a malicious threat agent can undermine the protective effect of diversity, because the threat is targeted instead of being fortuitous or accidental.
	Safety integrity levels	Safety standards such as [19] [20] prescribe different safety integrity levels for components of varying importance to safety, correlating these with the degree of rigour	Integrity levels are not commonly prescribed or included in regulations and there is correspondingly little regulatory

		required in the development and validation of these components.	guidance on where the security development effort should be concentrated.
	Security Information and Event Management Systems (SIEMs)	Existing safety standards such as [19] do not recommend the use of artificial intelligence for the highest criticality systems.	Sophisticated SIEMs and augmented intelligence analysis techniques – such as those recommended for use with critical national infrastructure [21] – rely on the use of AI.
	Development	Safety engineering requires that the changes to the development environments must be minimised where possible and an impact assessment performed for any changes [19] [22].	If it is known that a development environment will not be patched / updated for the length of time it takes to develop the system, this creates an opportunity for an attacker.
Design	Diversity and redundancy	Although avoidance of complex supply chains is encouraged, this is primarily to minimise difficulties in obtaining sufficient information about the safety properties of individual components, and is not prioritised over the principles of diversity.	A design which incorporates multiple diverse components complicates the supply chains and can leave the system vulnerable to cyber security attacks, because there are more potential points of vulnerability within the different supply chains.
	Communications and encryption	Safety engineering encourages the reduction of complexity systems under the ALARP principle. Encryption should therefore only be used where the consequent increase in complexity can be justified, both in terms of timing properties and potential failures of the encryption itself.	Encryption is a cornerstone of secure communications principles [23] and is encouraged for sensitive data without reference to a principle of proportionality.
	Timing and Power Analysis	Timing analysis in safety engineering has a significant focus on Worst Case Execution Time (WCET). Non-deterministic timing delays are therefore a potential safety concern.	Encryption algorithms including RSA are vulnerable to timing attacks [24], for which one potential mitigation is the use of a randomness algorithm to add a non-deterministic delay and hence prevent fixed-time computation.
	Timing and Power Analysis	Varying chip internal clock frequency to protect against power analysis attacks represents a change in the hardware properties of the system. Safety engineering principles therefore require an impact assessment of the change and a reflection of this in the safety case.	Power analysis attacks can be mitigated against by varying the chip internal clock frequency [25].
	Allow lists and authorization controls	Updating an allow list represents a change in the software properties of the system. Safety engineering principles therefore require an impact assessment of the change and a reflection of this in the safety case.	As part of the defence-in-depth layers of cyber security protection, allow lists are subject to the same ongoing expectation of updates as the rest of the system. Failure to update an allow list can result in a cyber security compromise.

Operation and maintenance	Maintenance	Alterations to a safety-critical system, including installation of patches and upgrades, carries a requirement to revalidate the impact of the new code to the degree of assurance consequent on the system's criticality.	Security requires continuous monitoring and updating of the system as required to respond to emerging threats. Installing patches and upgrades is an integral part of this procedure.
	Forensics and response	Quarantining or isolating a single system within a wider System of Systems (SoS) is not necessarily a fail-safe operation, and the focus is therefore on transitioning the wider SoS into a safe state. This may prevent immediate quarantine and analysis of the compromised system.	After a security incident, the procedure is focused on containment, which can typically involve quarantining a compromised system for analysis.
	Forensics and response	Safety culture encourages sharing information about accidents and near-misses with the community and public [26] [27].	Sharing information about an ongoing or emerging security incident can spread knowledge of a vulnerability and compromise other systems and, although encouraged [28], information sharing is to constraints regarding improper disclosure that can lead to adverse consequences.
	Forensics and response	Safety critical systems typically rely on a property of graceful degradation to prevent catastrophic failure, which property can be threatened by immediate system shut-down.	A potential response to a system under active cyber attack is to turn off the system and restore a back-up [1]. This is particularly prevalent in organisations which do not have robust standard operating procedures and response measures, and can compromise graceful degradation.
	Forensics and response	With the exception of near-misses and accidents resulting in long-term harm, failures of safety critical systems which result in accidents are typically obvious and immediate.	Security compromises and breaches can occur without any visible signs and may persist without detection.
	Forensics and response	Evolving safety incidents must be stopped or mitigated as soon as possible; there is no benefit to allowing harm to continue.	A cyber attack might be allowed to unfold in order to protect high valued assets that are not impacted or to collect incident response or threat intelligence data.
Risk assessment	System boundaries	Safety analysis techniques such as HAZOP and Failure Modes and Effects Analysis (FMEA) rely on a fixed and static knowledge of the system architecture and boundaries.	Cyber security assessments typically take into account the possibility of a change in system boundary, e.g. by connection of new equipment or introduction of malicious code.
	Risk attributes and estimates	Failure models for software assume static failures: the software will always fail in the same way. Historical data is therefore used in assessing achieved safety, to predict the likelihood of a failure of a given system under a given circumstance.	The likelihood of a security compromise fluctuates based on circumstances relevant to the attackers as well as the system, e.g. public knowledge of the vulnerability, availability of patches, time of day etc. Historical data may therefore not be given a primary significance.

	Risk attributes and estimates	The _risk = likelihood * consequence_ equation used to model risk assumes that likelihood of failure and consequence of that failure are independent.	The likelihood of attack by a threat agent is dependent on a number of factors, including the motivation of the agent. Higher consequences of an attack (e.g. significant financial gain) can increase the motivation of the agent. Likelihood and consequence can therefore not be considered independently
		UK law [3] requires that the safety risk posed by a system should be shown to be reduced As Low As Reasonably Practicable (ALARP), and carries a ‘proportionality’ expectation that risk reduction measures must be implemented unless it can be shown that the benefit gained from these measures is not proportionate to the reduction.	Cyber security guidelines and regulations do not carry an explicit requirement of proportionality.
	Risk attributes and estimates	Quantitative assessment is feasible for some aspects of safety analysis and standards allow not just for the calculation of risk but the calculation of confidence in that assessed risk.	Cyber security assessment depends significantly on human factors such as motivation, and hence security assessments typically do not assign quantitative values to these.

IV. SHARED CONCEPTS AND METHODOLOGIES

As discussed earlier, a number of existing works of literature attempt to describe and refine the tensions between safety and security [1] [2] [5] [8]. However, many of these are aimed at a relatively high level and are applicable to company goals, strategies and policies rather than development-level efforts and trade-offs.

In this section we present a discussion of cultural concepts which are shared between the safety and security domains, and identify how methodologies drawing on these concepts can be used to minimise conflicts as identified in Section 3. This discussion introduces ways in which developers can seek to fulfil the high-level guidance described above by the use of methodologies and cultural concepts which do not favour one domain over the other.

Defence-in-depth is a recommended practice within both the safety and cyber security domains [15] [16]. It relies on the provision of overlapping mitigations or layers of protection, so that a failure of any single one of these does not necessarily lead to failure of the entire system.

Safety critical systems make use of the defence-in-depth principle when assessing hazard mitigations. Acceptable mitigation strategies include eliminating the cause of the hazard entirely, breaking the link between cause and hazard such that the hazard itself is eliminated, reducing the likelihood that a hazard will progress to an accident, and reducing the severity of any accident which does result. Some or all of these strategies may be implemented, resulting in multiple layers of protection. This is of particular value where certain hazards are inherently associated with the system’s capability and cannot be eliminated.

Cyber security also promotes the defence-in-depth principle, although historically the focus has been on elimination of vulnerabilities rather than implementation of strategies to manage, understand and mitigate these vulnerabilities. This is rapidly being superseded, however, with the focus of modern security management and security assessment models being on understanding and managing vulnerabilities in the technology stack [29]. A 4-phase system for modelling opportunity and managing threats that draws on the defence-in-depth principle is presented in [30].

There are clear similarities between the way safety and security both implement the defence-in-depth principle. Crucially, this principle calls for a recognition that safety hazards and cyber security threats cannot feasibly be wholly eliminated: it is not possible to design a system that is either “completely safe” or “completely secure”. Because of this, development using defence-in-depth permits the integration of both safety and security concerns: a conflict that has been resolved by foregrounding the relevant safety (security) technique can be mitigated in the security (safety) domain by using an additional layer of protection to accommodate the hazard or vulnerability associated with this choice. This principle therefore allows developers to accept the presence of certain hazards or vulnerabilities without significantly compromising either safety or security properties.

V. Conclusion

Safety and cyber security are both fundamentally concerned with the integrity and dependability of systems, but within that concern may differ in terms of priorities, terminology, methodologies and culture. The distinction between the two has historically not been explicit, which has led to misinterpretation of the aims of the two domains and a lack of recognition of how conflicts and differing priorities may present.

We have introduced a development-level taxonomy of some of the most common conflicts between safety and security. We have also identified shared methodologies and cultural concepts, drawing on the principle of defence-in-depth, and proposed how an increased emphasis on multiple layers of safety and security provision could help address some of the conflicts identified. We propose in future work to develop this focus on defence-in-depth by means of a case study, examining the extent to which additional layers of protection can allow developers to accommodate both domains.

References

[1] National Academies of Science, Engineering & Medicine (2017), ‘Software update as a mechanism for resilience and security’, Proceedings of the Forum on Cyber Resilience Workshop Series, https://www.nap.edu/download/24833#

[2] Institute of Engineering and Technology (2021), ‘Code of Practice: Cyber Security and Safety’, https://electrical.theiet.org/guidance-codes-of-practice/publications-by-category/cyber-security/code-of-practice-cyber-security-and-safety/

[3] Health and Safety Executive (2001), ‘Reducing Risks: Protecting People’, https://www.hse.gov.uk/managing/theory/r2p2.pdf

[4] European Parliament and the Council of the European Union, (2016), ‘Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union’, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016L1148&from=EN

[5] Bloomfield, R., Netkachova, K. & Stroud, R. (2013), ‘Security-informed Safety: If It’s Not Secure, It’s Not Safe’, Proceedings of the International Workshop on Software Engineering for Resilient System, pp. 17 - 32.

[6] National Cyber Security Centre (2017), ‘TRITON Malware Targeting Safety Controllers’, https://www.ncsc.gov.uk/information/triton-malware-targeting-safety-controllers

[7] National Audit Office (2018), ‘Investigation: Wannacry Cyber Attack and the NHS’, HC 414, https://www.nao.org.uk/wp-content/uploads/2017/10/Investigation-WannaCry-cyber-attack-and-the-NHS.pdf

[8] Johnson, C. (2016) ‘Why We Cannot (Yet) Ensure the Cyber-Security of Safety-Critical Systems’, Proceedings of the 24th Safety Critical Systems Symposium, https://scsc.uk/scsc-131

[9] National Institute of Standards and Technology (2012), ‘Computer Security Incident Handling Guide’, NIST Special Publication 800-61, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r2.pdf

[10] Kelly, T. (2003), ‘A Systematic Approach to Safety Case Management’, SAE Technical Paper 2004-01-1779.

[11] Boyer A. et al. (2016) Modular Safety Assurance. In: Pohl K., Broy M., Daembkes H., Hönninger H. (eds) Advanced Model-Based Engineering of Embedded Systems. Springer, Cham. https://doi.org/10.1007/978-3-319-48003-9_10

[12] Williams, L. (2019), ‘Secure Software Lifecyle Knowledge Area’, National Cyber Security Centre, https://www.cybok.org/media/downloads/Secure_Software_Lifecycle_KA_-_Issue_1.0_August_2019.pdf

[13] Stevens, F. (2021) ‘Comparative Maritime Safety’, in Nawrot, J. & Peplowska-Dabrowska Z. (eds.) *Maritime Safety in Europe: A Comparative Approach*, Abingdon, Routledge

[14] Software Engineering Institute, Carnegie Mellon University (2016) ‘*SEI Cert C Coding Standard: Rules for Developing Safe, Reliable and Secure Systems’*, Carnegie Mellon.

[15] International Nuclear Safety Advisory Group (1996) ‘*Defence In Depth In Nuclear Safety’*, STI/PUB/013.

[16] Department of Homeland Security (2016) ‘*Recommended Practice: Improving Industrial Control System Cybersecurity With Defense-in-Depth Strategies’*, US Government.

[17] UK Government (2018), ‘The Network and Information System Regulations 2018’, https://www.legislation.gov.uk/uksi/2018/506/made

[18] Choudhary, N. (2018), ‘The Role of Safety Risk Management in the UK Rail Industry When Dealing With Cyber Threats’ in International Joural of Safety and Security Engineering, Vol 8, pp. 48 – 58

[19] International Electrotechnical Commission (2010), ‘Functional safety of electrical / electronic / programmable electronic safetyrelated systems’, IEC standard 61508, https://www.iec.ch/safety

[20] International Standards Organisation, (2011), ‘Road vehicles – functional safety’, ISO standard 26262, https://www.iso.org/standard/43464.html

[21] International Standard.Royal United Services Institute for Defence and Security Studies (2020), ‘Artificial Intelligence and UK National Security Policy Considerations’, https://rusi.org/sites/default/files/ai_national_security_final_web_version.pdf

[22] RTCA SC-205 (2011), ‘Software considerations in airborne systems and equipment certification’, RTCA/DO-178C, http://www.rtca.

[23] National Centre for Cyber Security, ‘Secure Communications Principles’, https://www.ncsc.gov.uk/guidance/secure-communication-principles-alpha-release

[24] Kocher, P. (1996) ‘Timing Attacks on Implementation of Diffie-Hellman, RSA, DSS and other systems’, in Advances in Cryptology, Lecture Notes on Computer Science 1109, Berlin, Springer-Verlag, pp. 104 – 113.

[25] Mayhew, M. and Muresan, R. (2016), ‘An overview of hardware-level statistical power analysis attack countermeasures’, Journal of Cryptographic Engineering, Vol 7, pp 213 – 244.

[26] Clarke, S. (1998), ‘Safety culture on the UK rail network’, Work & Stress Issue 3, pp 285 – 292.

[27] Wiegmann, D., Zhang, H., von Thaden, T., Sharma, G. & Gibbons, A. (2004) ‘Safety Culture: An Integrative Review’, The International Journal of Aviation Psychology, Vol 14, pp. 117-134

[28] Johnson, C., Badger, L., Waltermire, D., Snyder, J. & Skorupka, C. ‘Guide to Cyber Threat Information Sharing’, NIST 800-150, 2016, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-150.pdf

[29] Vidalis,S., Angelopoulou, O. (2013). “Deception and Manoeuvre Warfare Using Cloud Resources”, Journal of Information Security: a global perspective, volume 22, pp. 151 – 158.

[30] Morakis, E., Vidalis, S., Blyth, A. (2003), "Measuring Vulnerabilities and their Exploitation Cycle", Elsevier Information Security Technical Report, Vol 8, pp. 45 – 55