EVALUATING THE IMPACT OF CYBER SECURITY AND SAFETY WITH HUMAN FACTORS IN RAIL USING ATTACKER PERSONAS

Amna Altaf, BSc MSc, Bournemouth University, Fern Barrow, Poole, UK
Eylem Thron, BEng PhD, Ricardo Rail, 30 Eastbourne Terrace, London, UK
Shamal Faily, BSc DPhil, Bournemouth University, Fern Barrow, Poole, UK
Huseyin Dogan, BSc MSc EngD PGCert (HE), Bournemouth University, Fern Barrow, Poole, UK
Alexios Mylonas, BSc MSc PhD, Bournemouth University, Fern Barrow, Poole, UK

SUMMARY

Railway safety and security are typically considered as two independent engineering concepts, nonetheless it is now recognised that cyber security threats pose a risk to human life. Rail technology is engineered from a technical and safety perspective, and is subject to independent assurance. However, in the same way that a passenger or driver's view of a system is not the same as that of an assessor, attackers view rail systems and vulnerabilities in different ways. This raises the question of how cyber security can be better designed and assessed using an attacker-centric view of the system as part of the implementation of railway projects, while still accounting for safety and human factors.

To illustrate the impact of security on safety, we consider the example of a Polish Tram Incident in 2008, where a teenager converted a TV remote control into an infrared transmitter. This activated rail switches and redirected trams, leading to derailments and emergency stops which endangered the safety of passengers and railway staff. Although this well-known incident is used as an appeal to fear, it also highlights the need to remain resilient in the face of emerging threats and undesirable consequences.

In this paper, we present our approach for obtaining this resilience to leverage the relationship between human factors, safety concerns and security concepts using attacker personas. We also illustrate the use of the open-source Computer Aided Integration of Requirements and Information Security (CAIRIS) platform based on Integrating Requirements and Information Security (IRIS) framework for conceptualising the security and usability elements of the Polish Tram Incident, to identify the root cause safety and security problems related to human factors.

1 INTRODUCTION

The evolution of the rail infrastructure has highlighted cyber security concerns along with safety and human factors. Security might be compromised due to multiple vulnerabilities unseen by security engineers, but visible to attackers. To make the system and its associated environment safe and secure, we need to protect it against emerging threats and associated risks.

Both security and safety consider risk as a common factor. In broader terms, the malicious risk is defined as a security concern whereas the accidental risk is defined as a safety hazard. As such, safety and security are complementary (Kriaa *et al.*, 2015). Another commonality observed between safety and security relates to human computer interaction. For example, previous work (Ki-Aries and Faily, 2017) has identified a relationship between security breaches and possible human interactions, as both involved the system and environment.

Security engineers are responsible for anticipating possible threats and risks at design level, but taking an attacker's perspective helps in identifying exploitable system vulnerabilities. The attackers have the capabilities and motivations to dissect and leverage the possible weaknesses of the system which are left unattended by the security designers (Atzeni *et al.*, 2011).

One way of modelling the attacker centric view of the system is by building *Attacker Personas* based on a user-centered design approach (Steele et al., 2008). Personas are typically used to understand the behaviours of possible system users. Thus, the attacker personas represent the behaviour of archetypical attackers who are going to exploit the possible weaknesses of the system, causing a security breach leading to safety and human

factor implications. To illustrate their usefulness, we consider the real-life scenario of a Polish Tram Incident that took place in 2008.

The rest of the paper is structured as follows. In Section 2, we describe the related work upon which our approach is based, before presenting our approach in Section 3. We illustrate this approach with a case study example in Section 4. This is followed by discussion of attacker personas application in Section 5, before concluding with future directions for our work in Section 6.

2 RELATED WORK

2.1 Basics of Threat Modelling

To highlight the causes for possible incidents, including safety hazards (ISO/IEC 27002, 2007), identifying and categorising threats is a necessary part of risk analysis. This can be achieved by conceptualising various threat actors or agents involved within an environment and understand their interactions with the system.

Attack trees and misuse cases are the most well-known techniques used to model attacks during threat modelling. These approaches are used to understand the perspective of attackers (Schneier, 2000), but they need to be linked with some ground knowledge about attackers (Sindre and Opdahl, 2005).

Previous work has identified several factors that influence the malicious behaviour of threat actors and how the system vulnerabilities are exposed (Jones and Ashenden, 2005). Several categories of potential threat actors like IT experts, students, employees, hacktivist etc., have been defined by Open Web Application Security Project (OWASP, 2001). Similarly, the Common Attack Pattern Enumeration and Classification reflects the attack patterns that explain the system exploitation done by the attackers (CAPEC, 2007).

These system exploitation opportunities are known as vulnerabilities used by attackers, which lead to threats. The threats are associated with risks causing security breaches which might have hidden safety and human factors implications.

2.2 Personas and Attacker Personas

Personas model the archetypical behaviour of users. These are based on ground information collected from similar environments, where the users are expected to interact with some product or service being designed (Cooper, 1999). Norman (2004) states that the system design can be understood well from an assumptive perspective. For personas, the data sources and information obtained are backed up by imagining a variety of roles in which the personas are likely to be categorised (Pruitt and Grudin, 2003). The Encyclopaedia of Human-Computer Interaction, 2nd Ed. (2013) identifies four categories of personas:

- **Goal-directed:** The process and work flow that the user is going to perform in order to achieve its objective.
- **Role-based:** The user's role within an organisation based on both quantitative and qualitative data.
- **Fiction-based:** The assumptions made about the persona based on the experience or interaction of design team.
- **Engaging Personas:** A combination of goal-directed and role-based personas, giving a more detailed understanding.

Personas can be supported by stories and scenarios. A refined system view can be obtained by generating personas within relevant narrative scenarios and real-life situations (Nielsen, 2013). Moreover, story-based personas have a better chance of explaining the user behaviours (Pruitt and Grudin, 2003). This way the personas can be used to explain the different contexts of use they operate in. A persona built from a user-centered design approach also has a better chance of being used for various analysis purposes (Faily and Fléchais, 2010), e.g. threat modelling and risk analysis.

One way of achieving an attacker-centric view of the system is by building attacker personas (Shostack, 2014). The attacker personas visualise the problem space where there is a risk of a security compromise (Atzeni *et al.*,

2011). The possible threats faced by the system based on system vulnerabilities, which are otherwise not visible during security design phase can be discovered from an attacker's perspective.

3 APPROACH

We devised an approach for building attacker personas based on the Toulmin's Argumentation Model (Grounds, Warrants, Rebuttals) (Atzeni *et al.*, 2011). This approach entails qualitative data analysis and argumentation models to form the basis of personas characteristics. This approach is tool-supported by the Computer Aided Integration of Requirements and Information Security (CAIRIS) open-source platform [1](Faily, 2018). Lately, personas and CAIRIS have been used for case studies to determine the security and usability design issues (Faily *et al,.* 2015). In building attacker personas, we aim to better understand the safety and security implications of rail infrastructure along with human factors.

3.1 Data Source and Document References

The first step is the identification of appropriate data sources. These data sources are used for the purpose of building attacker personas, by laying the foundation for document references. These document references are the *factoids* stating major and minor details which contribute towards personas characteristics at later stages.

3.2 Affinity Diagramming

The document references (in the form of factoids) as elicited by carefully reviewing the data sources before carrying out affinity diagramming to make sense of the factoids. For this purpose, a *Trello*[2] board can be used to organise the factoids into different groups. Affinity diagramming entails organising the factoids into clusters sharing similar characteristics. Affinity diagramming is a participative activity, and would be carried out in one or more sessions with rail stakeholders.

3.3 Argumentation Models and Personas Characteristics

Based on the affinity groups established, the personas characteristics are defined. These are further categorised as *activities*, *attitudes*, *aptitudes*, *motivations*, or *skills*. The factoids from Trello board are imported into CAIRIS to define these personas characteristics based on document references. These document references are based on external documents which serve as backing source for personas characteristics.

CAIRIS allows these personas characteristics to be backed-up by building argumentation models. These argumentation models are based on Toulmin's Model of Argumentation (Toulmin, 2003). This entails categorising the document references within each affinity group as *grounds*, *warrants* or *rebuttals*. Grounds act as evidence for the validity of personas characteristics; warrants act as inference rules allowing the grounds to be linked with personas characteristics, and rebuttals act as counter-argument for the persona characteristics (Faily, 2018). These argumentation models provide a means to quickly validate each identified personas characteristic.

3.4 Attacker Personas Narrative

The attacker personas scenario-based narrative is written based on personas characteristics. The primary purpose of narrative is to describe the personas relationship with the system and the environment in which it is behaving.

4 CASE STUDY – POLISH TRAM INCIDENT

We evaluated our approach by applying it to an incident where a security breach occurred by exploiting a system vulnerability, which lead to a compromise of passenger safety. The 2008 incident was logged as *School Boy Hacks into Polish Tram System* in the `Repository of Industrial Security Incidents' as:

> *"A 14-year old boy, a Polish student, hacked into the tram system which enabled him to change track points in Lodz, Poland. Four trams were derailed. Twelve people were injured when a train derailed. The boy built an*

[1] https://cairis.org
[2] https://trello.com

infrared device that looked like a TV remote control that could control all the junctions on the line. No deaths occurred. The boy faced a special juvenile court on charges of endangering public safety (RISI, 2008)."

4.1 Data Source and Document References

The initial aim is to identify the possible threat actors in railway infrastructure, so the source data is based on real life incidents. Mainly the data is based on online e-articles written about the particular Polish Tram Incident helping us to notice each and every minor detail of the attacker from different perspectives. 47 factoids were document referenced from seven data sources as mentioned in Table 1.

Table 1: Online Articles used as Data Source for Building Attacker Personas

Ser.	***Article Title***	***Author***	***Publisher***
1.	Hacking Polish Trams	Bruce Schneier	Schneier on Security
2.	Polish Teen Derails Tram After Hacking Train Network	John Leyden	The Register
3.	Polish Teen Hacks His City Train, Chaos Ensues	Chuck Squatriglia	Wired Article
4.	School Boy Hacks into City's Tram System	Graeme Baker	The Telegraph
5.	Teen Derailed Trams with Home-made Device	Local Police	The Sydney Morning Herald Article
6.	School Boy Hacks into Polish Tram System		Repository of Industrial Security Incidents Log
7.	Teen Hacker in Poland Plays Trains and Derails City Tram System	Shelley Smith	Homeland Security

Based on these online articles and incident records, we concluded that attacker did not wish to intentionally cause harm. Instead, the attack was exploratory in nature with no consideration given to its consequences. The attacker also lacked the funding and resources to conduct the attack. Curiosity and passion were identified as the major motivation, and the attacker was equipped with no more than basic knowledge about the information and railway sector.

4.2 Affinity Diagramming

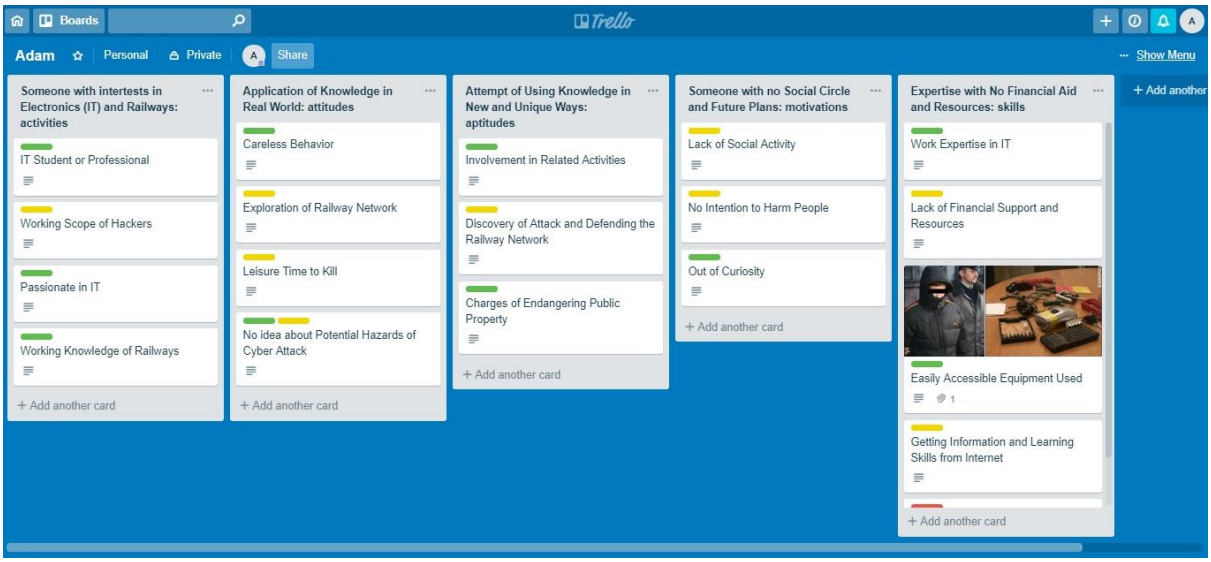

Figure 1: Affinity Diagramming using Trello

Using Trello board, the 47 factoids identified were arranged in different categories and clusters. Among which the most notable categories were personal characteristics, daily routines, interests, socialisation, scope of hacking,

ideas about consequences, motivations, careless attitude, work expertise, IT knowledge etc. Input from rail stakeholders helped in identifying the major and minor details related to incident.

4.3 Argumentation Models and Personas Characteristics

The factoids from Trello board were imported into CAIRIS and a persona named as *Adam* was created. Using CAIRIS, *Adam* was defined in environment *Morning Shift* and assigned the role of *Attacker*.

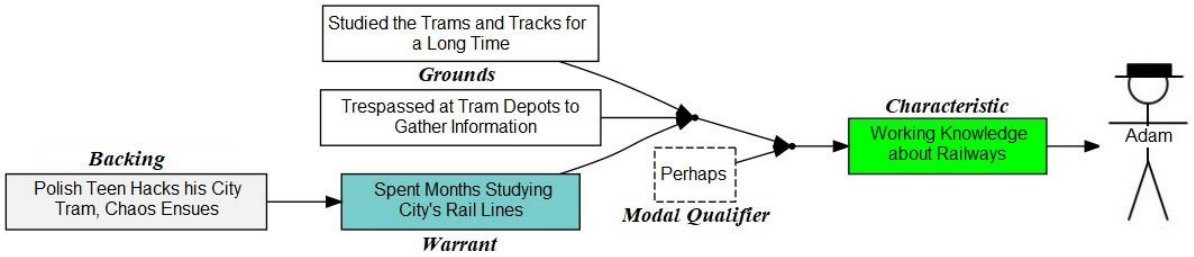

Figure 2: Argumentation Model for Personas Characteristic

Based on the affinity diagram, 18 argumentation models were defined, and used to specify 18 complementary persona characteristics.

4.4 Attacker Personas Narrative

The narrative for persona *Adam* was broken down into sections for Activities, Attitudes, Aptitudes, Motivations, and Skills. For each section, the persona characteristics were used to guide the section-specific narrative. For example, the narrative about the activities of *Adam* is as follows:

Adam is a teen-age IT student with good academic skills. Adam takes keen interest in his 'Electronics' class and is considered a genius by his teachers. After reading about railways during his school exercise, the boy has started to show some interest in railways. He has spent months studying about Lodz Tram System. He is often found killing his leisure time by trespassing at Tram Depots to gather information and equipment which can be re-used in creative ways. Apart from that, learning the coding skills from open-source public libraries by using the internet is also his favourite job.

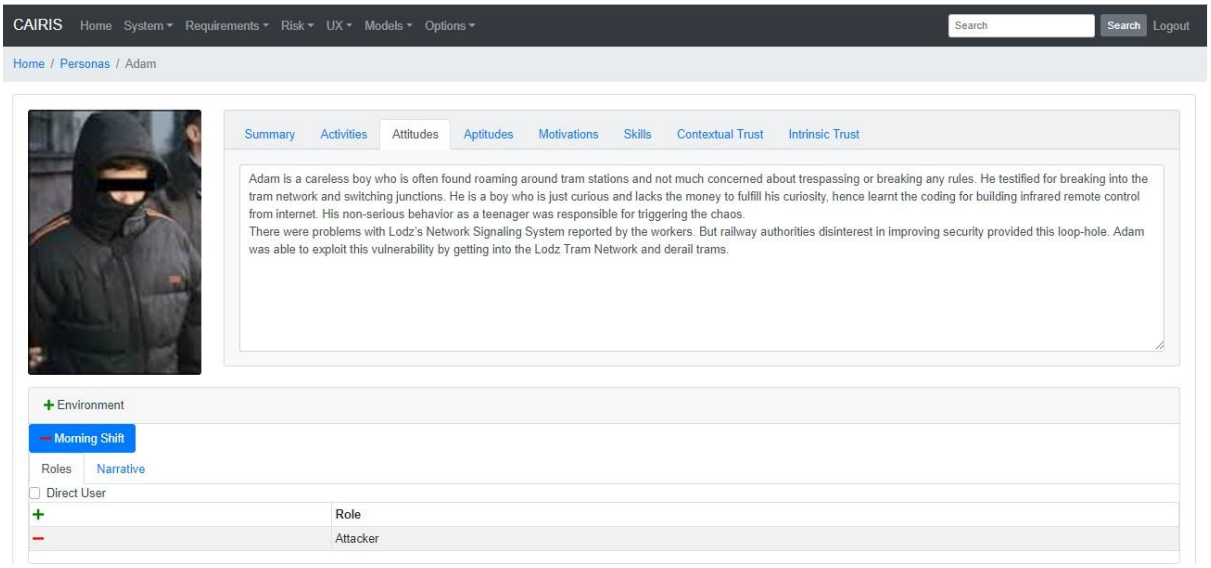

Figure 3: Attacker Personas Narrative using CAIRIS

5 DISCUSSION – APPLICATION OF ATTACKER PERSONAS

Based on the persona narrative, assumptions about the security, safety and human factor issues could be made. For instance, *Adam* appeared able to exploit the system vulnerability of *Faulty Track Points* based on his expertise in IT and coding skills which he learned from the Internet and during his class.

5.1 Cyber Security Concerns

Adam gives us an idea about the possible thinking of an attacker; thereby providing security engineers a better chance to identify the system vulnerabilities which can lead to threats. In this way, using the threat modelling features of CAIRIS, we can anticipate possible risks.

From the argumentation models and narrative for *Adam*, four major vulnerabilities are identified namely, *faulty track points*, *reported problems with signalling system*, *1970s switching system*, and *lack of risk assessment*. These vulnerabilities lead to the identification of three threats namely, *switch splitting*, *replay attack*, and *network intrusion*. These threats formed the basis for risk modelling in CAIRIS and lead to the recognition of following three risks: *train derailment*, *unauthorised access into signalling system* and *injury of railway staff or passengers*.

5.2 Potential Safety Hazards

Sometimes emergent threats impact environmental safety along with system security. Using attacker personas, this overlap between safety and security was better understood. For example, *Adam* was able to cause a security breach through which he intruded the railway network and switched tram directions, subsequently injuring multiple people. Fatalities were avoided only due to the timely intervention by authorities. Here *disruption of services*, *accidental collision between two or more than two trains* and *loss of life of staff or passengers* are identified as the potential safety hazards faced due to the security breach.

5.3 Human Factor Issues

Attacker personas help visualise the possible tasks scenarios. The parameters of these task scenarios -- specifically task duration, frequency, and user demands -- were determined using CAIRIS. Moreover, the tasks further helped identify the system level goals and user level goals. This provided a means for better visualising the system's behaviour and drawing links between security and usability (human factors) in the form of goal-obstacle and responsibility modelling. For example, Adam *learned the coding skills* from his class and the Internet before he *built an infrared device* by modifying a universal remote control. Adam used that infrared device to *record signals and replayed* them to *switch track points*. The completion of these tasks leads to the satisfaction of system goals (*Modify TV Remote Control*, *Access Railway Network* and *Redirect Railway Trams*) on the part of attacker. The task scenarios and hence the attacker personas can also be fed into the training needs analysis to identify the safety, security and human factors training needs to reduce risks and improve performance.

6 CONCLUSION

With reference to the rail industry, this paper has shown how attacker personas can be built, with the assistance of CAIRIS as tool-support. The modelling of the Polish Tram Incident scenario using attacker personas, in the form of risk, task, goal-obstacle and responsibility was shown to determine the inter-relationship between safety, security and human factors engineering.

By modelling the Polish Tram Incident Using attacker personas, the hidden vulnerabilities in the system responsible for causing harm to the broader environment were identified. These vulnerabilities lead to the recognition of threats, which informed risk analysis, and, on the basis of this analysis, the associated safety hazards like *disruption of services*, *accidental collision between two or more than two trains* and *loss of life of staff or passengers* were determined. The personas narrative also formed the basis for the identification of human factors concerns like tasks, responsibility and goal-obstacle modelling.

However, the validation of this persona is left as future work. By involving the right stakeholders from railway sector, the attacker personas can be further refined. Even the attacker personas for different types and categories of threats actors like nation-state actors, organised crime, hacktivist, inside actors etc, can be conceptualised to better understand the cyber-security and safety concerns along with human factors. Moreover, these personas can be used to model tasks and goal-obstacles of the system, along with threat and risk modelling. Using these different models, different views of system like environment, asset, task, goal, risk and responsibility can be presented. From these different views the security and usability design decisions can be made as early as possible, and the associated safety hazards are likely to become visible as well.

7 ACKNOWLEDGEMENTS

We are grateful to Daniel Woodland, Alex Bishop and Alzbeta Helienek for their valuable input on this paper.

8 REFERENCES

Atzeni, A. *et al.* (2011) 'Here's Johnny: A Methodology for Developing Attacker Personas', in *2011 Sixth International Conference on Availability, Reliability and Security*. *2011 Sixth International Conference on Availability, Reliability and Security (ARES)*, Vienna, Austria: IEEE, pp. 722–727. doi: 10.1109/ARES.2011.115.

Schneier, B. (2000) 'Secrets and Lies: Digital Security in a Networked World', in. John Wiley & Sons.

'Common Attack Pattern Enumeration and Classification' (2007). Available at: http://capec.mitre.org/.

Cooper, A. (1999) *The Inmates Are Running the Asylum*. Macmillan Publishing Co.

Faily, S. (2018) *Designing Usable and Secure Software with IRIS and CAIRIS*. Cham: Springer International Publishing. doi: 10.1007/978-3-319-75493-2.

Faily, S. and Fléchais, I. (2010) 'Barry is not the weakest link: Eliciting Secure System Requirements with Personas', p. 8.

Faily, S., Lyle, J., Fléchais, I., and Simpson, A. 'Usability and Security by Design: A Case Study in Research and Development', in *2015 Proceedings of the NDSS Workshop on Usable Security, Internet Society*.

ISO/IEC (2007) 'ISO/IEC 27002: Information Technology – Security Techniques – Code of Practice for Information Security Management.' ISO/IEC.

Jones, A. and Ashenden, D. (2005) 'Risk management for computer security : Protecting your network and information asset', in. Butterworth-Heinemann.

Ki-Aries, D. and Faily, S. (2017) 'Persona-centred information security awareness', *Computers & Security*, 70, pp. 663–674. doi: 10.1016/j.cose.2017.08.001.

Kriaa, S. *et al.* (2015) 'A survey of approaches combining safety and security for industrial control systems', *Reliability Engineering & System Safety*, 139, pp. 156–178. doi: 10.1016/j.ress.2015.02.008.

Nielsen, L. (2013) *Personas - User Focused Design*. London: Springer-Verlag (Human–Computer Interaction Series). Available at: https://www.springer.com/gb/book/9781447159032 (Accessed: 6 March 2019).

Norman, D. (2004) *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books.

Pruitt, J. and Grudin, J. (2003) 'Personas: Practice and Theory', in *Proceedings of the 2003 Conference on Designing for User Experiences*. New York, NY, USA: ACM (DUX '03), pp. 1–15. doi: 10.1145/997078.997089.

RISI - The Repository of Industrial Security Incidents (2008). Available at: https://www.risidata.com/ (Accessed: 19 November 2018).

Shostack, A. (2014) *Threat modeling: Designing for Security*. Indianapolis, IN: John Wiley and Sons. Available at: http://site.ebrary.com/lib/bournemouth/Doc?id=10837601 (Accessed: 21 November 2018).

Sindre, G. and Opdahl, A. L. (2005) 'Eliciting security requirements with misuse cases', *Requirements Engineering*, 10(1), pp. 34–44. doi: 10.1007/s00766-004-0194-4.

Steele, Adam & Jia and Xiaoping (2008) 'Adversary Centered Design: Threat Modeling Using Anti-Scenarios, Anti-Use Cases and Anti-Personas', *Proceedings of the 2008 International Conference on Information & Knowledge Engineering, IKE 2008, July 14-17, 2008, Las Vegas, Nevada, USA*, pp. 367–370.

The Encyclopedia of Human-Computer Interaction, 2nd Ed. (2013). Available at: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed (Accessed: 6 March 2019).

'The Open Web Application Security Project' (2001). Available at: http://www.owasp.org/index.php/.

Toulmin, S. (2003). 'The Uses of Argument'. Updated Edition. Cambridge University Press