Identifying safety and human factors issues in rail using IRIS and CAIRIS.

ALTAF, A., FAILY, S., DOGAN, H., MYLONAS, A. and THRON, E.

2020

Identifying Safety and Human Factors Issues in Rail using IRIS and CAIRIS

Amna Altaf[1], Shamal Faily[1], Huseyin Dogan[1], Alexios Mylonas[1], and Eylem Thron[2]

[1] Bournemouth University, Fern Barrow, Poole, UK
{aaltaf, sfaily, hdogan, amylonas}@bournemouth.ac.uk
[2] Ricardo Rail, 30 Eastbourne Terrace, London, UK
eylem.thron@ricardo.com

Abstract. Security, safety and human factors engineering techniques are largely disconnected although the concepts are interlinked. We present a tool-supported approach based on the Integrating Requirements and Information Security (IRIS) framework using Computer Aided Integration of Requirements and Information Security (CAIRIS) platform to identify the safety and human factors issues in rail. We illustrate this approach with a case study, which provides a vehicle for increasing the existing collaboration between engineers in security, safety and human factors.

Keywords: Security by Design; Safety Hazards; Human Factors; IRIS.

1 Introduction

As the rail information infrastructure becomes integrated with operational technology, new vulnerabilities are introduced together with the new threats that exploit them. As such attacks are directly or indirectly responsible for compromising safety, cyber security has become a new concern for rail safety engineers. Poor design decisions made during security engineering may lead operators to make human errors or mistakes where rules are intentionally disobeyed [14], which may eventually affect system safety. Therefore, rail infrastructures can only be made safe and secure if along with safety and security, the human factors engineers contribute to its design and evaluation.

In this paper, we illustrate such an approach where the core concepts from the Integrating Requirements and Information Security (IRIS) framework are used to define an intersecting model, based on a proposed relationship between different security-by-design and usability techniques. This approach is tool-supported using the open-source Computer Aided Integration of Requirements and Information Security (CAIRIS) platform[3].

A key contribution of this work is the use of Human Factors Analysis and Classification System (HFACS) to augment IRIS framework and CAIRIS platform to identify safety and security issues. This helps rail stakeholders better

[3] https://cairis.org

understand the safety and human factors implications of security concerns, and also helps discover inter-dependencies between security, safety and human factors engineering techniques.

In Section 2, we describe the related work upon which our approach is based, followed by the explanation of our approach in Section 3. We illustrate this approach with a case study example in Section 4, before concluding and discussing future directions for our work in Section 5.

2 Related Work

2.1 Security and Safety Challenges in Rail Infrastructure

The rail infrastructure has long been managed in accordance with health and safety standards, working within legislative requirements such as in United Kingdom the Railway Act 2005, under guidance and supervision from bodies like Railway Safety and Standards Board (RSSB) and Office of Rail Regulation (ORR). More recently, the shift to digitalisation stipulated by the European Railway Traffic Management System (ERTMS) imposed by European Union has seen the introduction of the Common Safety Method for Risk Evaluation and Assessment (CSM-REA) in addition to UK specific safety concepts such as 'As Low As Reasonably Practicable' (ALARP) in managing safety risks.

The evolving nature of the cyber threats have imposed a greater challenge for security experts in rail [12]. As a result, the rail infrastructure needs to be supported by codes of practice (CoPs) throughout its life cycle as a combination of security and safety [6]. Security should be infused with safety at a design phase by ensuring a combined risk assessment approach.

Similarly, the strong linkage between the human intent to violate rules and imposed safety hazards described by [3] highlights the value of combining safety with human factors. The Human Factors Analysis and Classification System (HFACS) is a framework for eliciting possible accident and incident contribution factors based on taxonomy of active and latent failures caused by human interactions in rail [17]. The HFACS have been used by rail stakeholders to determine the human error sources behind accidents and incidents. However, to date, there has been no work on how it can be used to consider safety or security attributes of rail system.

2.2 Bridging Security, Safety, and Human Factors

Hazards and accidents may occur due to security breaches, and dependability – delivering services that can justifiably trusted – encompasses safety and some major elements of security [5]. Safety is an attribute of dependability, with availability, reliability, integrity and maintainability; security refers to the availability and integrity attributes and to confidentiality [13]. Thus the risk factors (probability of chances of damage) along with the dependability (trust and reliance on system) are triggered by safety and security issues. Both safety and security

engineering communities are now working to better bridge their communities [11], e.g. safety engineering consideration of *security mindedness* [6].

Previous work has considered human error as an intersecting concept between cyber security and safety. Humans may cause harm by making mistakes (active failures) or by inducing errors within system (latent failures) [7], with human intent as a differentiating factor. If humans are benevolent (unintentional), they may alert the safety engineers by causing hazards and accidents; if malevolent (intentional), they may carry out threats and exploit vulnerabilities that compromise system security [16], thereby leading to a risk instigating a safety hazard.

2.3 IRIS and CAIRIS

The Integrating Requirements and Information Security (IRIS) process framework [8] was devised to understand how design concepts associated with security, usability, and software engineering could be aligned. It is complemented by the Computer Aided Integration of Requirements and Information Security (CAIRIS) platform, which acts as an exemplar for tool-support to manage and analyse design data collected when applying an IRIS process. IRIS and CAIRIS have been used in several real-world case studies, including the development of security policies for critical infrastructure systems [9].

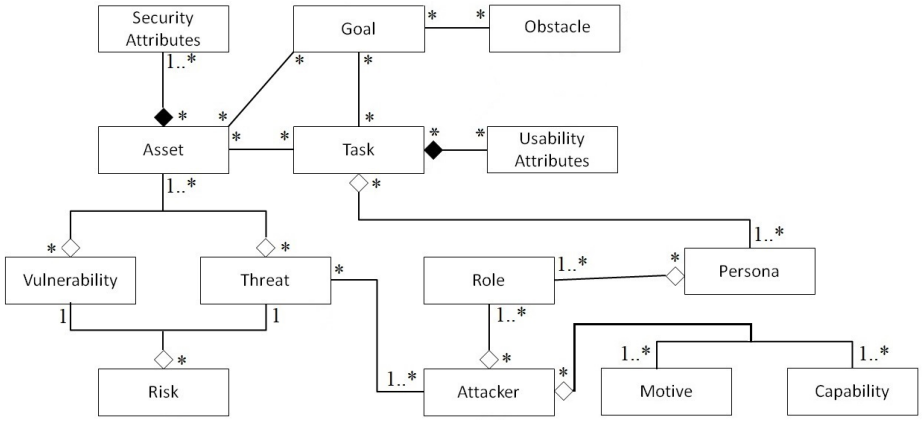

Fig. 1. UML Class Diagram of IRIS

The core IRIS concepts are illustrated in the UML class diagram in Fig. 1. Vulnerabilities and threats contribute to potential risks, and threats are contingent on attacker's intent. This intent helps analysts identify the tasks and goals they carry out or exploit, which can help determine human factors issues in the form of human errors (active failures). Consequently, although not explicitly de-

signed with safety in mind, IRIS provides a foundation for integrating security, safety and human factors.

3 Approach

We have devised an approach based on the IRIS framework, which leverages security and usability engineering approaches to better understand the safety implications of rail infrastructure under design. This approach is tool-supported using CAIRIS. The approach takes input from security and human factors engineers, as well as from rail stakeholders with safety expertise.

3.1 Asset Modelling and their Associations

The approach begins with a security analysis of the system and its environment by identifying the possible assets [10]. These assets and their relationships are modelled using UML class diagrams. Each asset is defined in a particular environment, and categorised by asset types. The security attributes for assets like confidentiality, integrity, availability are defined and values (Low, Medium, High) are assigned, based on priorities defined by the rail stakeholders.

3.2 Roles and Attacker Personas

The roles are defined based on stakeholder roles in rail like driver, manager, ticketing staff, signaller etc. The roles are further used to identify specific personas describing the archetypical behaviour of system actors. Attacker personas are created by following the approach described in [4]; this approach entails using qualitative data analysis and argumentation models to form the basis of personas characteristics. *Factoids* underpinning the personas are elicited by categorising data about attackers, and thematically analysing these factoids based on affinity groups. CAIRIS facilitates online affinity diagramming, and allows annotated factoid lists to be imported into CAIRIS as personas characteristic argumentation models. These argumentation models are based on Toulmin's model of argumentation, such that each characteristic is justified by one or more *grounds* that evidence the persona's validity, *warrants* that act as inference rules connecting the grounds to the characteristic, and *rebuttals* that act as counterarguments for the characteristic. A model qualifier is also used to describe the confidence in the validity of the personas characteristic. Attacker personas narratives are then specified based on these personas characteristics.

3.3 Vulnerabilities Identification and Threat Modelling

The vulnerabilities are weaknesses of the system, which, if exploited, leads to a security breach [8]. While identifying vulnerabilities, the assets open to attack are identified. Personas support this exercise by providing an insight into an attacker's mind, given that an attacker's model of the system may be different

from a security engineer's model of the same system. Attacker's motivation and capabilities play an important role in threat identification. Tasks and goals fulfilled by attackers also provide an insight during threat modelling. The threats identified are assigned security properties based on the goals of attacker.

3.4 Risk Analysis

Vulnerabilities and threats contribute to the identification of potential risks [8]. Using risk analysis, the likelihood and severity of an incident is determined based on the ability of an attacker, and the value of assets that need to be protected. CAIRIS generates visual risk models based on this analysis, which are used as the basis for further analysis.

3.5 Task and Goal-Obstacle Modelling

Based on asset modelling and risk analysis, the concerned tasks and goals are elicited. These form the basis of system and user level goals. Tasks and goals are identified from the attacker's perspective and also form the basis for *obstacles* that model obstructions to system goals. Goal and task models can help the security engineers to better understand the system threat model.

3.6 Identification of Safety Hazards

The risk model generated by CAIRIS determines the safety hazards, by showing the linkage between the assets with their associated security attributes, vulnerabilities, emergent threats and the possible risks. The main purpose of this type of modelling is to identify the possible safeguards to be taken and minimise the chances of occurrence of any hazardous events.

3.7 Human Factors Analysis

Our approach uses HFACS as a multi-level framework defining human factors in four main categories [15]: unsafe acts of operations, preconditions for unsafe acts, unsafe supervision and organizational influences. In lieu of a standardised methodology for determining the human error sources using HFACS, each vulnerability, threat and risk identified as part of threat model is analysed against the human factors definitions according to HFACS. The value with the closest possible explanation for human error is labelled as the desired human factors issue.

4 Case Study - Polish Tram Incident

We illustrate our approach by applying it to a real life incident where a security breach occurred by exploiting a system vulnerability, leading to the compromise

of passenger safety[4]. The 2008 incident was logged as *School Boy Hacks into Polish Tram System* in the 'Repository of Industrial Security Incidents' [2].

We gathered open source intelligence as an input to our approach. This was based on several online articles written about the particular Polish Tram Incident. We supplemented publicly available data with the Operational Concept for European Railway Traffic Management System (ERTMS); this was used to understand the system architecture, application levels, operating modes, signalling principles and control. We also obtained feedback on the emerging CAIRIS model from safety and human factors experts at Ricardo, who were representative of the rail stakeholder that might provide input to our approach.

4.1 Asset Modelling and their Associations

Two working environments were defined: *Morning* and *Night shift*. The *Morning Shift* is based on assumption that it is expected to be much busier in terms of passenger numbers, compared to operations that take place during *Night Shift*. 51 assets were identified, based on types of software, hardware, information and people. Assets were modelled by taking an attacker's perspective of the tram system, thus helping the security engineers to understand the relevant vulnerabilities. Asset modelling was not limited to the early stages of the process; at later stages asset associations were also defined. For instance, during attacker personas definition three assets namely *Infrared Remote Control*, *Public Libraries* and *Internet Codes* were identified. These assets formed the basis for determining the capabilities of an attacker who learned the coding for building infrared remote control from the Internet.

4.2 Roles and Attacker Personas

The analysis about the rail infrastructure lead to the recognition of 11 roles. The most notable was the role of *Attacker*. Based on online articles and incident records, we concluded that the attacker did not wish to intentionally cause harm. Instead, attacks were exploratory in nature with no consideration given to the consequences. The role of attacker further motivated us to understand the intent and capability behind the cyber attack with the help of personas.

We created an attacker persona *Adam* based on relevant sources for the Polish Tram Incident, which provided different perspectives of the incident. *Adam* was built based on 18 argumentation models used to specify 18 complementary personas characteristics, underpinned by 47 factoids. For example, the persona characteristic *Working Knowledge about Railways* describes how *Adam* gained access to the rail network based on his skills and knowledge; he recorded and replayed signals using a universal remote control. Based on this, we identified a system vulnerability, i.e., the *1970s Switching System* on which Poland Tram System was operating, and the subsequent threat of *Unauthorised Access into Poland Railway Signalling System*.

[4] The final model created, including references to online sources used, is available from: https://bit.ly/2KSocEg

4.3 Vulnerabilities Identification and Threat Modelling

By exploring the attacker motives, 4 vulnerabilities were identified namely, *Poor Architectural Design and Lack of Risk Assessment*, *1970s Switching System*, *Reported Problems with Signalling System* and *Faulty Track Points*. These vulnerabilities were responsible for compromising the security of 6 assets.

We also identified 3 threats: *Poland Railway Network Intrusion*, *Replay Attack* and *Switch Splitting*. The anticipation of possible threats and cyber-attacks at design level is the work of security engineers, but considering *Adam's* perspective helped identify exploitable vulnerabilities. For example, the threat *Poland Railway Network Intrusion* was based on our interpretation of *Adam's* ability to exploit *Faulty Track Points*.

4.4 Risk Analysis

Within an environment of *Morning Shift*, 4 risks were defined using vulnerabilities and threats. These form the basis of the risk analysis, the results of which are illustrated in Fig. 2.

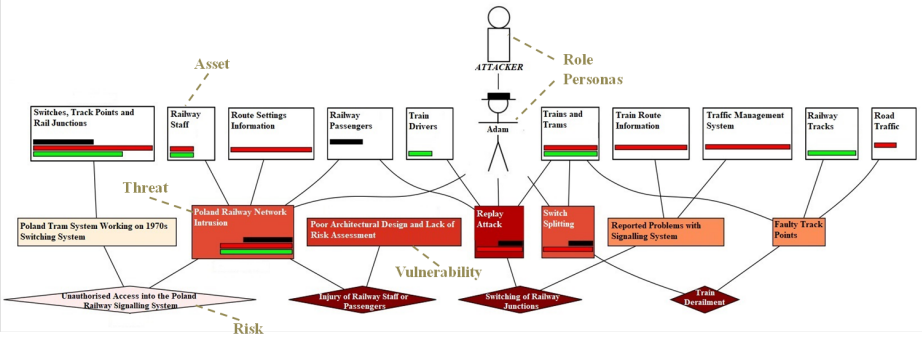

Fig. 2. Risk Modelling in CAIRIS

The threat of *Switch Splitting* based on vulnerability of *Faulty Track Points*, could lead to risk of *Train Derailment*. On the basis of this risk, security design decisions that minimise the chances of occurrence of this risk can be taken. The risk analysis also contributed towards the better understanding of visible safety hazards and human factors issues based on their occurrence and likelihood ratios.

4.5 Task and Goal-Obstacle Modelling

The narrative of attacker personas formed the basis for responsibility modelling which comprised of identification of 4 tasks performed by attacker to conduct the cyber-attack. Adam *learned 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 lead to the satisfaction of system goals (*Modify TV Remote Control*, *Access Railway Network* and *Redirect Railway Trams*) on the part of attacker.

The attack was conducted by exploiting system loop-holes. The exploitation of these loop-holes were active failures on the part of security engineers. For example, the vulnerability *Reported Problems with Signalling System* led to the human factors issue of *Violations* as the operation and performance of signalling system was not compliant with secured protocols and standards. This allowed the attacker to perform the task of *Record Signals*, fulfilling the system goal *Access Railway Network*. In this case, the major security goal defined by security engineers which would have acted as an obstacle for attacker would have been the use of *Advanced Train Control Protocol System* which would have denied *Adam* an unauthorised access into the railway network. Thus, it would have mitigated the cyber-attack, and ensured the safety of passengers.

4.6 Identification of Safety Hazards

For explanation purposes, we consider the risk of *Switching of Railways Junctions* that is due to the threat of *Replay Attacks*. The realisation of this risk might cause *Collisions* between two or even more than two trains, which compromises the safety of passengers and staff present in train. Table 1 represents the identification of potential safety hazards from risk modelling elements (vulnerabilities, threats, risks) based on the Risk Assessment Log presented by Randstad Rail [1]. The documentation of Randstad Rail includes the activities and tasks in the railway sector which may lead to catastrophic hazards. The identified risks were used to categorise these safety hazards. Knowledge of these potential safety hazards is helpful for alerting safety engineers dealing with critical infrastructures.

4.7 Human Factors Analysis

Table. 1 shows how the vulnerabilities, threats and risks identified can be categorised to determine the human factors issues based on HFACS along with safety hazards. These human factors issues also help us to verify the system usability for risks, by the satisfaction of user goals depending on certain procedures, competencies, permissions and training needs analysis (TNA) to achieve those goals and complete defined tasks.

For example, the risk of *Injury of Railway Staff or Passenger* which is linked to threat of *Poland Railway Network Intrusion*, may lead to safety hazard of *Loss of Life*. In this case, the human factors issue observed using the HFACS framework is the poor design of *Technological Environment* due to *Poor Architectural Design and Lack of Risk Assessment*, which has life-threatening consequences. This illustrates how the timely evaluation of technological environment using checklists and task factors can minimise the chances of risk occurrence.

Table 1. Human Factors Issues based on HFACS

Vulnerabilities	Threats	Associated Risks	Safety Hazards	Human Factors
Faulty Track Points	Switch Splitting	Train Derailment	Life Threatening for Staff and Passengers in Train as well as near Train	Failed to Correct Known Problems
Reported Problems with Signalling System	Replay Attack	Switching of Railway Junctions	Collision (Between Two Trains or Even More Than Two Trains)	Violations
Poland Tram System Working on 1970s Switching System	Poland Railway Network Intrusion Threat	Unauthorised Access into the Poland Railway Signalling System	Disruption of Train Services or Emergency Stop	Inadequate Supervision
Poor Architectural Design and Lack of Risk Assessment	Poland Railway Network Intrusion Threat	Injury of Railway Staff or Passengers	Loss of Life	Technological Environment

5 Discussion and Conclusion

In this paper, we presented a tool-support approach for identifying safety and human factor issues, based on core concepts from IRIS and CAIRIS. The scientific novelty has been the methodological application to safety and human factors engineering in rail. We carried out a preliminary evaluation of this approach by applying it to a case study where inter-dependencies between safety, security, and human factors were present. In doing so, we have made three contributions. First, our approach shows how asset modelling and their associations, can be used to identify security attributes namely, confidentiality, integrity, availability of assets as prioritised by rail stakeholders. Second, we have shown how building models of attackers not only rationalises attacker assumptions, but also helps to identify system vulnerabilities. Both lead to the identification of threats which, with the support of scenarios, rationalises risks and the identification of several safety hazards. On the basis of these hazards, root causes of active failures (human errors) like *violations* and *inadequate supervision* could be determined using HFACS. Finally, we have shown how building the personas for other roles like driver and signaller helps rail stakeholders determine the task scenarios in more detail. These task scenarios can be used by human factors engineers to inform hierarchical and cognitive task analysis which can predict the reliability of systems in different environments.

We are evaluating our approach on a project where the representative rail stakeholders will be closely involved when considering the risks, roles, tasks, goals, requirements, dependencies and obstacles between the humans and systems. In future work, we will present a refined process-framework based on best practices from safety, security and human factors engineering. For this purpose, further categorisation of tasks at system, design or operator levels using ERTMS specifications may have the potential to determine broader design weaknesses. A more thorough task analysis exercise could provide more detailed insights into human factors, and subsequent security and safety concerns. The resultant process-framework will be translated into tool-support for implementation in rail and other critical infrastructures.

Acknowledgements

The work described in this paper was funded by the BU studentship *Integrating Safety, Security, and Human Factors Engineering in Rail Infrastructure Design & Evaluation*. We are also grateful to Ricardo for their support.

References

1. Randstad Rail - Generic Risk Assessment Log. https://www.randstad.co.uk
2. RISI - The Repository of Industrial Security Incidents. https://www.risidata.com (2008)
3. Alper, S.J., Karsh, B.T.: A systematic review of safety violations in industry. Accident Analysis & Prevention **41**(4), 739–754 (Jul 2009)
4. Atzeni, A., Cameroni, C., Faily, S., Lyle, J., Flechais, I.: Here's Johnny: A Methodology for Developing Attacker Personas. In: 2011 Sixth International Conference on Availability, Reliability and Security. pp. 722–727. IEEE, Vienna, Austria (Aug 2011)
5. Avizienis, A., Laprie, J.C., Randell, B., Landwehr, C.: Basic concepts and taxonomy of dependable and secure computing. IEEE Transactions on Dependable and Secure Computing **1**(1), 11–33 (Jan 2004)
6. Bloomfield, R., Bishop, P., Butler, E., Stroud, R.: Security-Informed Safety: Supporting Stakeholders with Codes of Practice. Computer **51**(8), 60–65 (Aug 2018)
7. Brostoff, S., Sasse, M.A.: Safe and Sound: A Safety-Critical Approach to Security p. 10 (2001)
8. Faily, S.: Designing Usable and Secure Software with IRIS and CAIRIS. Springer International Publishing, Cham (2018)
9. Faily, S., Flechais, I.: User-Centered Information Security Policy Development in a Post-Stuxnet World. In: 2011 Sixth International Conference on Availability, Reliability and Security. pp. 716–721. IEEE, Vienna, Austria (Aug 2011)
10. Gollmann, D.: Computer Security. Wiley & Sons, 2nd edn. (2007)
11. Jonsson, E., Olovsson, T.: On the Integration of Security and Dependability in Computer Systems p. 6 (1998)
12. niv. Lille Nord de France, F-59000 Lille, French Institute of Science and Technology for Transport, Development, and Networks IFSTTAR-COSYS-ESTAS, Villeneuve d'Ascq, France, Boudi, Z., Koursi, E.M.E., Ghazel, M.: The New Challenges of Rail Security. Journal of Traffic and Logistics Engineering (2016)
13. Piètre-Cambacédès, L., Bouissou, M.: Cross-fertilization between safety and security engineering. Reliability Engineering & System Safety **110**, 110–126 (Feb 2013)
14. Reason, J.: Human Error. Cambridge University Press (1990)
15. Wiegmann, D.A., Shappell, S.A.: A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System. Routledge, Aldershot, Hants, England ; Burlington, VT, 1 edition edn. (Jul 2003)
16. Young, W., Leveson, N.G.: An integrated approach to safety and security based on systems theory. Communications of the ACM **57**(2), 31–35 (Feb 2014)
17. Zhou, J.L., Lei, Y.: Paths between latent and active errors: Analysis of 407 railway accidents/incidents' causes in China. Safety Science **110**, 47–58 (Dec 2018)