# Integrated design framework for facilitating systems-theoretic process analysis.

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

2022

# Integrated Design Framework for Facilitating Systems-Theoretic Process Analysis

Amna Altaf[1], Shamal Faily[2], Huseyin Dogan[1], Eylem Thron[3], and Alexios Mylonas[4]

[1] Bournemouth University, Poole, UK
{aaltaf,hdogan}@bournemouth.ac.uk
[2] Robert Gordon University, Aberdeen, UK
s.faily@rgu.ac.uk
[3] CCD Design & Ergonomics Ltd, London, UK
eylem.thron@designbyccd.com
[4] University of Hertfordshire, Hatfield, UK
a.mylonas@herts.ac.uk

**Abstract.** Systems-Theoretic Process Analysis (STPA) helps mitigate identified safety hazards leading to unfortunate situations. Usually, a systematic step-by-step approach is followed by safety experts irrespective of any software based tool-support, but identified hazards should be associated with security risks and human factors issues. In this paper, a design framework using Integrating Requirements and Information Security (IRIS) and open source Computer Aided Integration of Requirements and Information Security (CAIRIS) tool-support is used to facilitate the application of STPA. Our design framework lays the foundation for resolving safety, security and human factors issues for critical infrastructures. We have illustrated this approach with a case study based on real life *Cambrian Coast Line Railway* incident.

**Keywords:** STPA; Safety Hazards; Security Risks; Human Factors; IRIS; CAIRIS; Rail Infrastructure.

## 1 Introduction

Systems-Theoretic Process Analysis (STPA) is used to identify control actions and causal factors behind accidents to improve system design [21]. The approach revolves around a series of pre-defined steps followed by experts. Using STPA analysis, the identified safety hazards can also mitigate security risks. For example, poor design decisions may lead operators to make human errors or mistakes where rules are un-intentionally disobeyed [19]. Consequently, the system safety and security may be compromised due to human intervention in the form of errors or violatioins.

Integrating Requirements and Information Security (IRIS) framework has been used to identify security risks leading to safety hazards for identifying human factors issues [3]. This is achieved by identifying and modelling assets

associations, roles and personas, vulnerabilities, threats, risks, tasks and goals [13]. Based on the IRIS framework and complementary Computer Aided Integration of Requirements and Information Security (CAIRIS) platform, assumptions about security concerns and human factors issues are explicated for critical infrastructures. The framework allows complementary human factors approaches to be used to derive use case specifications based task analysis modelling to determine human failure levels leading to errors or mistakes [4]. These failure levels are used to identify associated safety and security design solutions by identifying potential hazards.

An extended design framework can be formulated by integrating these human factors and security methods for facilitating safety analysis using STPA. By conducting STPA using the IRIS framework and CAIRIS platform. This aims to resolve safety, security and human factors design concerns for critical infrastructures. To demonstrate this approach, we have used the real life incident of *Cambrian Railway*. This case study serves as a guide for human factors, safety and security experts to deal with human factors issues, associated safety hazards and potential security risks.

The rest of the paper is structured as follows. Section 2 describes the related work and Section 3 describes our proposed design framework. Our design framework is demonstrated by applying it for case study in Section 4. This is followed by discussion and conclusion for future directions of our work in Section 5.

## 2 Related Work

### 2.1 Security and Safety Engineering

There are commonalities between safety and security engineering, with both communities now working to bridge their gaps [17]. Safety engineering can be considered from a security mindset [10], and the International Electro-technical Commission (IEC) has suggested a framework TC 65/AHG 1 for coordinating safety and security together [16].

Several existing approaches in safety and security engineering are complementary due to inter-linked concepts. The Defence-in-Depth (DiD) approach, which is also applied in security, was derived from a safety design of nuclear plants [27]. In security, the graphical representation of attacks related to attackers using attack trees was derived from fault trees for safety of systems [30]. A Hazard and Operability Study (HAZOP) is a structured and systematic approach used to identify and evaluate risk problems in safety. The concept has been applied to security because of risk dealing with security properties (confidentiality, integrity, availability) was discovered as a linking factor [36]. Similarly, Failure Mode and Effects Analysis (FMEA) approach from safety has been applied in security as Intrusion Modes and Effects Analysis (IMEA) [7].

**Systems-Theoretic Process Analysis** A consistent design approach for safety and security can be based on identifying safety hazards using Systems-Theoretic

Process Analysis (STPA) [37]. STPA is a safety hazard analysis process model for identifying control actions for possible hazards and accidents in causal scenarios [21]. STPA is derived from Systems-Theoretic Accident Model and Processes (STAMP) process model. STAMP revolves around examining components which operate independently and together by playing their part in a system. The accident causal models are derived by studying patterns and investigating accidents from a safety engineering perspective. The processes and components when interacting with each other give rise to safety and security emergent properties. The control actions and feedback required for controlling these emergent properties based on algorithms leads to recognition of controllers. These control actions and controllers (processes) are subsequently mapped. During design, these activities are considered as high-level functional safety requirements for system. An incorrect process model may lead to an accident, where four types of unsafe control actions may occur; these control actions may occur too soon, too late, incorrect or altogether are missing. This is also known as identification of causal scenarios for unsafe control actions [20].

Safety experts should consider security along with safety as part of STPA [26]; the cyber security considerations in STPA are expanded into the STPA-Sec development method for safety critical systems [25]. Using STPA-Sec, system and component level requirements are dissected to identify safety constraints. These safety constraints help identify hazard scenarios leading to violations. These violations are weaknesses or vulnerabilities in system that allow the loss (accident) to happen [34]. Usually, hazards may also be based on human and system interactions, especially human error [22] which is not acknowledged by STPA-Sec.

The UK's National Cyber Security Centre has introduced the application of STAMP/ STPA in various case studies for improving risk framework for cyber security problems. The cyber security risk toolbox have been modified to include STPA approach for enterprise IT infrastructure including automated/ connected products, industrial control systems and critical national infrastructure [5]. These case studies are used to inform about safety and security requirements in a socio-technical environment by considering the human involvement. These requirements further motivate the consideration of human factors for identifying human error source as an impacting factor behind cyber security.

STPA can potentially be used to identify human factors issues as a result of interactions with system, such as human error sources from human behaviour, and the labelling design flaws along with system hazard analysis. The unsafe behaviours behind system automation could be used to connect causal scenarios with hazard analysis. The causal scenarios helps to generate a series of possibilities with cause and effect relationship as a result of human interaction with system. Furthermore, this argument has been supported by applying this approach for case study of Automated Parking Driving System [15].

### 2.2 Safety and Human Factors Engineering

Human safety in critical infrastructures like rail is sometimes compromised due to the occurrence of human error [8, 24], so its identification during the design of safety critical systems should be a priority. The rail standard EN 50126-1 emphasises the consideration of human factors during rail system's design process along with Reliability, Availability, Maintainability and Safety (RAMS) [1]. Additionally, the risk assessment for design of safety of systems like transportation industry prescribes the use of a Human Reliability Analysis (HRA) approach [18].

Based on the Swiss Cheese Model of accident causation [29], multiple layers of defence exist within a system or an organisation to protect against emergent errors or mistakes that lead to accidents. The model takes the inspiration from a slice of cheese where the holes represent the human weaknesses and different slices act as the barriers. Some holes are active failures whereas some are latent failures; all holes must be aligned at the same time for the accident to occur. Latent failures originates from active failures and usually have same catastrophic effects on human life [29]. Due to the complexity of consequences of incidents, there is no well-defined methodology for determining the sources of these failures [32]. The *human* is the most important aspect of this model, whose intent and capabilities are typically variable. Therefore, not all possible holes can be generalised before time. Based on Reason's error taxonomy [29] of cognitive, behavioural, personal and organisational factors, the Human Factors Analysis and Classification System (HFACS) framework represents four levels of failures and error sources [35].

**Task Analysis Approach** Tasks are performed by users to achieve goals. These are assumptions made about the behavioural specifications of users involved and how they are supposed to interact with the system [12]. Task Analysis (TA) determines the set of tasks to be performed by users under observation. The TA is conducted by identifying the task for analysis, determining the associated sub-tasks and writing a step-by-step narrative for sequence of actions to be performed [2]. Previous work has shown how User Experience (UX) techniques can be used to conduct TA, using a combination of Cognitive Task Analysis (CTA) and Hierarchical Task Analysis (HTA) [4]. CTA identifies different types and values of cognitive reactions, which influence human performance during completion of tasks. HTA identifies task dependencies and sequences as a hierarchy, where high-level use cases are refined into low-level use cases. Using the use-case specifications format, different levels of human failures are then identified using tool-support [4].

### 2.3 Human Factors and Security Engineering

The threat to a system in an environment is usually caused by an attacker: the human element responsible for compromising the security [31]. This identifies humans as the biggest source for human error [29]. Similarly, the security engineers

now prioritise the human dimension of system during design phases by considering the usability attributes during asset identification, threat scenario, misuse case, task duration, responsibility modelling etc [14]. Therefore, the concept of effective information security revolves around the idea of Human Computer Interaction - Security (HCI-security) of the system using a user-centered approach [33].

**Integrating Requirements and Information Security** The Integrating Requirements and Information Security (IRIS) process framework [13] 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.

Using IRIS, vulnerabilities and threats contribute to potential risks, and threats are contingent on attacker's intent [3]. CAIRIS facilitates the creation of personas – narratives of archetypal users that embody their goals and expectations [23] – and the online data analysis that contributes to the specification of their characteristics as argumentation models [14]. Personas narratives are specified based on these characteristics, and supported by the narratives, analysts can identify the tasks and goals using the Knowledge Acquisition in autOmated Specification (KAOS) goal modelling language [11]). Collectively, these help determine human factors issues in the form of human errors (active failures). Personas narrative also contribute towards understanding capability, intent, action and motivation for stakeholder roles, and goal and task models help the security engineers better understand the system threat model on the basis of *obstacles* that obstruct to system goals. CAIRIS also helps to model use-cases and information assets as Data Flow Diagrams (DFDs) where different trust boundaries display various levels of privilege operating within system. Consequently, although not explicitly designed with safety in mind, IRIS and CAIRIS provides a foundation for integrating safety, security and human factors engineering.

## 3 Approach

Our design framework comprises of human factors informed safety analysis and security engineering. The human factors approach draws on the identification of roles, persona building, and the generation of task models and use-case specifications to apply a partial-STPA assessment. The process begins by identifying an accident or loss, where an unplanned situation during performance of tasks by specified roles or use-case actors may lead to catastrophic consequences. The safety engineers work to minimise these occurrences by incorporating safety checks and goals in system design whereas a security engineer focuses on vulnerability and threat recognition for risk analysis. Using CAIRIS, STPA models

include a KAOS goal model to show goals and obstacles contributing to the scenario behind the accident.

**Pre-requisite** Before applying STPA, the stakeholder roles are defined within system. The roles are further used to identify specific personas describing the archetypical behaviour of system actors. Personas are created by following the approach described by [6]. Persona narrative play a significant role in determining the actors intent and capabilities which contribute towards understanding task. Using personas narrative, the concerned tasks within imagined scenarios are elicited based on roles. These elicited tasks form the basis of system and user level goals. Tasks are defined as narrative text, with additional details on their dependencies, consequences, and benefits. The narrative helps to understand the objective of task along with its procedural description, but the persona plays a major role behind the recognition of tasks.

Using CAIRIS, a *Task Participation Form* relates personas with task using usability attributes such as duration, frequency, demands and goal conflict. The usability attributes with different values highlight tasks with different colours during task models. These task models comprise of tasks against specified roles and personas which facilitate the specification for use case actors and use cases for human factors analysis. These models also help relate associated assets, threats and vulnerabilities, which assist experts during security analysis.

With the help of personas narrative and task models, use case specifications are defined. Each use case specification comes with an objective, actor, pre-conditions, steps (task sequence), post-conditions and exceptions. The use-case actors can also be linked with task models, showing relationship between role, persona, task and use-case. These elaborate task models help experts to visualise design of system along with specified environment by conducting TA using use-case specification format [4].

**Step 1: Accident, Hazard and Constraint** The STPA process begins by defining the accidents (losses) in relation to identified hazards [21]. The system-level constraints are also defined at this stage. During TA, the tasks with *High* level of human failures are analysed for identifying accident (loss) and hazard. Using CAIRIS, the goal and obstacle modelling in KAOS captures accident, hazard and constraints. The *obstacle* with the type "loss" is used to model accident whereas type "hazard" models associated hazard. The constraints are modelled as *goal*. The visual representation of these linked concepts provide more meaning and understanding for further analysis by domain experts.

**Step 2: Model Control Structure** At this stage, a control structure of the major components and controllers within system, along with the commands used between them is sketched. The commands between components and controllers are usually labelled as control or feedback [21]. An effective way for modelling these control structures within CAIRIS is by using DFD. Using DFDs, the trust

boundary may variate between controller, controlled process, sensor or actuator. The processes and data stores are defined using use cases and information assets, and CAIRIS automatically visualises a control structure model as a DFD.

**Step 3: Unsafe Control Action** The worst case scenarios leading to hazards are recognised by defining unsafe control actions. An unsafe control action is a control action which is either applied too early or too late. The safety constraints are determined for minimising these unsafe control actions [21]. In CAIRIS, an unsafe control action is presented using *obstacle* and the safety constraint is modelled by associating these obstacles with DFDs.

**Step 4: Causal Factor** The causal factors are identified by analysing the controllers, processes, feedback and control paths [21]. In CAIRIS, the identified tasks during human factors analysis, are linked-up with hazards and system-level constraints using KAOS goal refinement associations. Here, the task model and personas narrative might also contain the detail for an occurrence of event known as causal factor. The model generated is known as the controller process model, which highlights the design-level issues leading to accident scenarios as a result of hazard. By using these models vulnerability, threat and risk analysis can help resolve security, safety and human factors design issues.

**Step 5: Risk Analysis Model** These identified causal factors are also defined as system vulnerabilities leading to hazards (accidents). The vulnerabilities are also system weaknesses, which, if exploited by attackers as threats, contribute to the realisation of risks. The core IRIS concepts are used for modelling risk elements in the form of attacker, threat and vulnerability. The assets and their associations already defined during STPA are used in this risk analysis. 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. Threat scenarios (misuse cases) are also defined to evaluate the rating of each risk. CAIRIS generates visual risk models based on this analysis, which are used as the basis of further security analysis.

## 4 Case Study - Cambrian Incident Investigation

The real life incident of *Cambrian Railway* is used to conduct a case study based on qualitative evaluation of presented design framework[5]. The incident took place in October 2017 on the Cambrian Coast Line in Wales, where a train oversped due to technical failure [9]. The train was following the route of Cambrian Coast Line. During service between Barmouth and Llanaber, the

[5] The final model created, including references to online sources used, is available at GitHub repository: https://github.com/s5121191/CyberICPS_21. This relies on the CAIRIS fork at https://github.com/s5121191/cairis.

train travelled at three times of its normal speed. The over-speeding was timely observed by its train driver, who immediately reported the fault to concerned authority. Following this, manual routing was conducted by the train driver and signaller until the fault was rectified. No accidents occurred and no human was harmed during this incident. A formal investigation was conducted by Rail Accident Investigation Branch (RAIB) and five recommendations were suggested to Network Rail [28].

We chose this incident based on multiple factors like signalling system, service type, form of rail transit, and design implementation. The Cambrian Coast Line implemented the *European Railway Traffic Management System (ERTMS)*. ERTMS is based on European Train Control System (ETCS) as a rail signalling system, which ensures reliability, optimised capability and automation. Achievement of these qualities in ERTMS depends on safe, secure and usable design goals. The service type is *Passenger Train*, which is safety critical, and the goal is to ensure safety and security of human life. The *Light Rail* is preferred as the form of rail transit because of rapid speed, inter-city passenger travel (familiarity of routes) and usable design features.

The Cambrian Incident[6] case study application of the integrated design framework begun with data collection. All open source (online) documentation and literature was collected and surveyed. Moreover, the relevant stakeholders were determined. This included safety expert for STPA process support, security expert for understanding causal factors (including risk analysis) and human factors expert for advise during goal-obstacle modelling, task and personas scenarios. For this project, two environments were identified namely, *peak* and *off-peak hours*. The *Peak Hours* were defined from Monday-Friday 0630-0930 and 1600-1900 hours, whereas the *Off-Peak Hours* were from Monday-Friday at all other times (minus Peak Hours) including all day on Weekends and Bank Holidays.

The Cambrian Incident case study was modelled using KAOS to show a general scenario behind the accident [28]. For this purpose, 6 goals and 4 obstacles were identified and their associations were defined as shown in Fig. 1, where different shades of obstacles were due to varying probability of occurrence; the darker the shade, the higher the probability. The model stated the major goal of *Auto Signalling Computer Restart* being obstructed by obstacle of *No Indication of an Abnormal IT Condition*. This goal was associated with sub-goal of *Temporary Speed Restriction (TSR) Data Uploaded*, where the obstruction was caused due to *Missing Independent Check*. The TSR data was displayed on Driver Machine Interface (DMI) available to train drivers. Therefore, come the sub-goal of *DMI Used for Operational Control Display*, this goal had two sub-goals defined along with an obstacle where *Speed Restriction Not Uploaded* caused a problem during its goal fulfilment. The sub-goal when *Fourth Passenger Train Service Operated* lead to obstacle where normal service delivery was compromised because of *2J03 Passed TSR from 30km per hour to 80km per hour*. This fault

[6] This case study is applied for demonstration purpose only and in no way undermines any previous findings or studies.

was timely reported by train driver to the IT technicians. Therefore, the goal of *Reported Fault on Train 2J03 Service* was fulfilled.

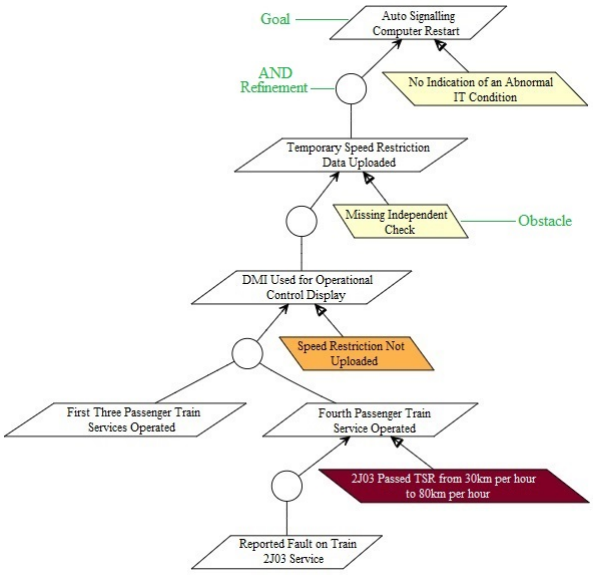


**Fig. 1.** Goal-Obstacle Model for Cambrian Incident Case Study

**Pre-requisite** The train driver and signaller roles were important in this incident. The train driver identified and reported the fault, then reverted to manual routing in order to ensure safety of passengers and normal service delivery. Alongside, the signaller was responsible for doing an independent check of upload of correct TSR. Upon recognition of fault, signaller reported it to technician and co-ordinated routes with train driver for no disruption of service.

Using CAIRIS, a total of 5 roles were identified including *on-board staff, on-board passenger, signaller, train driver* and *train maintainer*. Two personas, *Ray* and *Neil*, were created for the role of train driver and signaller respectively. Ray was based on 22 argumentation models. Neil's persona was based on 18 argumentation models. These argumentation models were used to understand persona characteristics, which formed the narrative for personas. This narrative and underpinning data analysis contributed to the identification of task models for further analysis.

A total of 19 tasks were created in CAIRIS; 11 were derived from Ray, and 8 from Neil. For example, the task of *Perform ETCS Self-Test Function* was found from persona characteristic of activities for Ray as shown by the bold text in the scenario below.

*Ray as train driver begins his job, by booking on and getting updated information on his laptop. This is based on documentation received about booking depot and preparing train for service.* ***Also, before operating train Ray is going to perform an on-board European Train Control System (ETCS) self-test function for finding faults and failures. He is going to produce a failure report and proceed only if the status of train for service is Safe and Fit.***

With the help of personas narrative and task models, 17 use-case specifications were defined.

**Step 1: Accident, Hazard and Constraint** During TA, 3 use cases *Combining Workstations, Granting Off-Peak Blockage* and *Conflict Prediction and Resolution* corresponded with *High* levels of human failure. Using these tasks, the accidents were defined using *obstacle* with type loss. In the given scenario 2 accidents were defined as *Collision Between Two or More Trains* and *Train Derailment*. The former was due to loss of operational control data for controlling trains and a cause of concern for road traffic, on-board passengers, staff, train driver and other trains. The latter occurred due to over-speeding where along with on-board passengers, staff, and train driver other concerns included were like movement authority signals, DMI, TSR and driver advisory information.

This was followed by recognition of 4 hazards with respect to these identified accidents, where each hazard was responsible for specified concerns in the form of assets. For example, the hazard of *Train Enters Uncontrolled State* was dependent on occurrence of accident of *Train Derailment*.

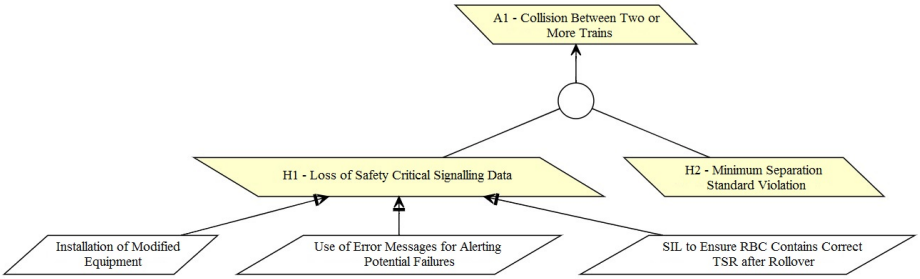


**Fig. 2.** KAOS Association Between Accident, Hazard and Constraint

At this point the constraints were modelled as goals. There were 8 constraints for preventing these hazards. For example, the hazard of *Loss of Safety Critical Signalling Data* had 3 constraints identified as *Installation of Modified Equipment*, *Use of Error Messages for Alerting Potential Failures* and *Safety Integrity Level (SIL) to Ensure Radio Block Center (RBC) Contains Correct TSR after Rollover* as shown in Fig. 2.

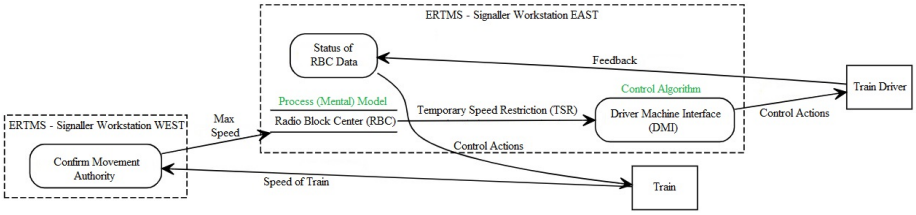


**Fig. 3.** DFD of Control Structure Model using CAIRIS

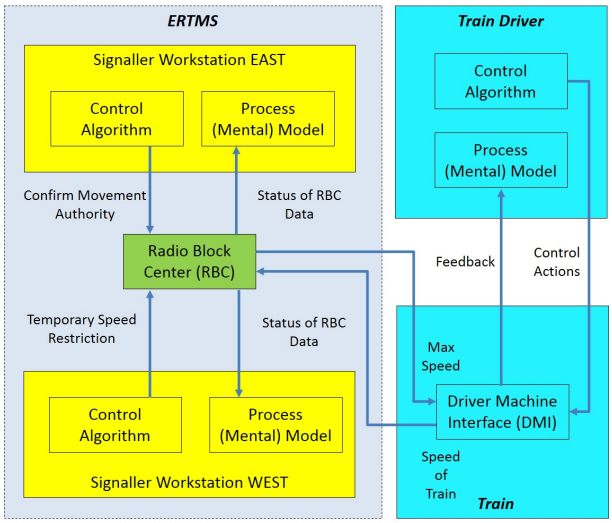


**Fig. 4.** High-level Control Structure Model

**Step 2: Model Control Structure** Using 17 use-cases and 29 information assets, the control structure was modelled. In CAIRIS the DFD for this case study consisted of three main elements: ERTMS, Train Driver and Train, where the flow of information between each element was taking place in order to display flow of control between processes as shown in Fig. 4. For example, behind the DFD element of *Train Driver* there are control actions and feedback of information flowing between control algorithms of *Driver Machine Interface* and *Status of RBC Data*. The DFD in CAIRIS, shown in Fig. 3, was also used to construct high-level control structure model as shown in Fig. 4.

**Step 3: Unsafe Control Action** Using UCA keyword, the unsafe control actions were defined in CAIRIS as obstacles. *UCA1 - ETCS Failure* and *UCA2*

- *Reliance on Procedures to Ensure TSR Application* were defined as *2* UCAs for this incident. UCA1 was related to ERTMS signalling control system and due to safety issues. UCA2 was related to RBC and occurred during RBC rollover. Using KAOS, these UCAs were linked to hazards. Therefore, the hazard of *Train Enters Uncontrolled State* was related to UCA1 and *Minimum Separation Standard Violation* was related to UCA2.

**Table 1.** Unsafe Control Action corresponding to Accident, Hazard and Constraint

| ***Accident (Loss)*** | ***Hazard*** | ***Constraint*** | ***Unsafe Control Action*** |
|---|---|---|---|
| A1 - Collision Between Two or More Trains | H1 - Loss of Safety Critical Signalling Data | Installation of Modified Equipment | Reliance on Procedures to Ensure TSR Application |
| | | Use of Error Messages for Alerting Potential Failures | |
| | | SIL to Ensure RBC Contains Correct TSR after Rollover | |
| | H2 - Minimum Separation Standard Violation | Implement a Mandatory Safety Assurance Procedure | ETCS Failure |
| A2 - Train Derailment | H3 - Trains Enter Uncontrolled State | Inclusion of defensive Programming (SQL) to Protect Against Unsafe State | ETCS Failure |
| | | Good Safety Management Engineering | |
| | H4 - Operational Planning Violation | Capture and Retention of Data for Investigating Failures | Reliance on Procedures to Ensure TSR Application |
| | | Robust Configuration Management | |

**Step 4: Causal Factor** At this stage, the identified tasks within human factors analysis were associated with constraints (goals). The model generated was known as the controller process model, where the tasks carry an explanation for unsafe control actions. For example, the constraint defined as *Implement a Mandatory Safety Assurance Procedure* was complemented by a task known as *Send Movement Authority*. The delay or incorrect *Movement Authority* had catastrophic consequences.

**Step 5: Risk Analysis Model** Using causal factors, risk modelling elements in the form of attacker, threat and vulnerability were also found. An hypothetical attacker was someone defined with capabilities such as knowledge, education and training of software and technology, with a motivation to breach system. 2 vulnerabilities with configuration type and critical severity were identified as *Lack of Safety Integrity Level* and *No Error Messages for Alerting Potential Failures*. Using these vulnerabilities, *2* electronic and malware type of threats were found namely, *Threat of ERTMS Safety Related Failure* and *Threat of Loss of Data Packets*. Each threat was assigned assets and valued for security properties including confidentiality, integrity and availability.

Consequently, these vulnerabilities and threats contributed to 2 risks with misuse cases as *Risk of Loss of Life due to Train Collision or Derailment* and *Risk of Failure of Signalling Network over ERTMS* as shown in Fig. 5. In the risk model, the elements were filled with different colours based on values of security properties, threat and vulnerability type and risk scoring. Like obstacles, the darker the shade, the more likely, severe, and impactful is the threat, vulnerability, and risk respectively.

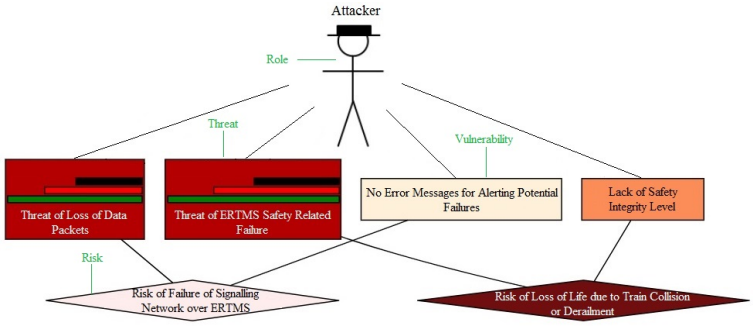


**Fig. 5.** Risk Model Based on Attacker, Threat and Vulnerability

## 5 Discussion and Conclusion

In this paper, STPA process model was derived using the IRIS framework and CAIRIS platform. As a result, three signification contributions are made. First, we demonstrate how the STPA process model is aligned with IRIS and CAIRIS, providing a single platform for all elements and contributing factors related to hazard analysis. These elements comprise of accident (loss), hazard, system constraint, component (control algorithm), process (mental) model, unsafe control action (obstacle) leading to causal factors. Second, we show how the causal factors including tasks can identify vulnerabilities, threats and risks present within system. This can be visualised using a security risk analysis model in CAIRIS. The risk model enlists tasks related to roles and personas which can be further analysed for use case specifications based task analysis as a combination of CTA and HTA leading to human error sources unlike STPA-Sec. Furthermore, the human error sources has the tendency to contribute towards potential safety hazards. Finally, the approach focused on bringing security and human factors methods support to STPA. Initially, the STPA process model is suggested by keeping in mind the safety where several case study applications suggested the involvement of human element. This human element is considerable in a socio-technical environment, where the system weaknesses (vulnerabilities) are highlighted by recognising human error sources. These human error sources establish grounds for understanding potential hazard scenarios and model better risk analysis. Hence, this research builds the scope of connection and integration between safety, security and human factors.

Using this integrated design framework, safety goals (safety constraints), security risks and human factors concerns (levels of human error) are highlighted. The STPA process model is derived from human factors approach which contribute towards the identification of potential safety hazards. These safety hazards are then used for identifying control actions and causal factors behind accidents for improving system design. The IRIS framework concepts alignment with STPA lead to better outcome as human perspective (task model and analysis) is understood in more detail. The risk model arising from STPA analysis

facilitates security experts as well. Moreover, by using CAIRIS, the effort required by safety, security and human factors experts is minimised by providing automated and efficient design solutions. These efficient design solutions enable experts from different domains to accomplish different tasks by combined and reduced effort.

For demonstration purposes, STPA method is applied using the case study of *Cambrian Incident*. The human factors approach such as identification of roles and personas, task analysis and use-cases are used to understand processes, asset associations and goal-obstacle models. In return, KAOS models and DFDs (processes and datastores) are used to apply STPA, where risk analysis based on recognition of attackers, threats, vulnerabilities, risks and misuse cases are done simultaneously. This helps to evaluate an integration of concepts between safety and security, security and human factors, and human factors and safety. This lays the foundation for overlapping concepts between three domains.

As future work, the application of safe, secure and usable design framework will be done on an industrial live project. For this purpose, safety, security and human factors experts will be consulted for validation of data and process behind approach.

## 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.*

## References

1. CENELEC - EN 50126-1 - Railway Applications - The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS) - Part 1: Generic RAMS Process — Engineering360. https://standards.globalspec.com/std/10262901/cenelec-en-50126-1 (Oct 2017)
2. Affairs, A.S.f.P.: Task Analysis. /how-to-and-tools/methods/task-analysis.html (Sep 2013)
3. Altaf, A., Faily, S., Dogan, H., Mylonas, A., Thron, E.: Identifying Safety and Human Factors Issues in Rail using IRIS and CAIRIS. In: CyberICPS 2019: 5th Workshop On The Security Of Industrial Control Systems & Of Cyber-Physical Systems. Springer, Luxembourg, Luxembourg (Sep 2019)
4. Altaf, A., Faily, S., Dogan, H., Mylonas, A., Thron, E.: Use-Case Informed Task Analysis for Secure and Usable Design Solutions in Rail. In: The 16th International Conference on Critical Information Infrastructures Security. Springer, Lausanne, Switzerland (Sep 2021)
5. Anna, G.: Methodological Findings from Applying STPA in Cyber Security Case Studies. In: MIT STAMP Conference. MIT Partnership for Systems Approaches to Safety and Security (PSASS), MIT Campus, Cambridge, USA (Mar 2019)
6. 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)
7. Babeshko, E., Kharchenko, V., Gorbenko, A.: Applying F(I)MEA-technique for SCADA-Based Industrial Control Systems Dependability Assessment and Ensuring. 2008 Third International Conference on Dependability of Computer Systems DepCoS-RELCOMEX (2008)
8. Baysari, M.T., McIntosh, A.S., Wilson, J.R.: Understanding the human factors contribution to railway accidents and incidents in Australia. Accident Analysis and Prevention **40**(5), 1750–1757 (Sep 2008)
9. BBC: 'Lessons learnt' over train speeding on Cambrian line. BBC News (Dec 2019)
10. Bloomfield, R., Bishop, P., Butler, E., Stroud, R.: Security-Informed Safety: Supporting Stakeholders with Codes of Practice. Computer **51**(8), 60–65 (Aug 2018)
11. Dardenne, A., van Lamsweerde, A., Fickas, S.: Goal-directed requirements acquisition. Science of Computer Programming **20**(1), 3–50 (Apr 1993)
12. Diaper, D., Stanton, N.: The Handbook of Task Analysis for Human-Computer Interaction. CRC Press, Mahwah, NJ (2004)
13. Faily, S.: Designing Usable and Secure Software with IRIS and CAIRIS. Springer International Publishing, Cham (2018)
14. Faily, S., Fléchais, I.: Barry is not the weakest link: Eliciting Secure System Requirements with Personas p. 8 (Sep 2010)
15. France, M.E.: Engineering for Humans : A New Extension to STPA. Thesis, Massachusetts Institute of Technology (2017)
16. IEC: IEC - TC 65/AHG. https://www.iec.ch (2019)
17. Jonsson, E., Olovsson, T.: On the Integration of Security and Dependability in Computer Systems p. 6 (1998)
18. Kirwan, B.: Validation of human reliability assessment techniques: Part 1 — Validation issues. Safety Science **27**(1), 25–41 (Oct 1997)
19. Lahoz, C.H.N.: Systematic review on STPA A preliminary study p. 34 (2015)
20. Leveson, N.: Engineering a Safer and More Secure World p. 72 (Jun 2011)
21. Leveson, N.: Systems-Theoretic Process Analysis Handbook p. 188 (Mar 2018)
22. Mindermann, K., Riedel, F., Abdulkhaleq, A., Stach, C., Wagner, S.: Exploratory Study of the Privacy Extension for System Theoretic Process Analysis (STPA-Priv) to Elicit Privacy Risks in eHealth. In: 2017 IEEE 25th International Requirements Engineering Conference Workshops (REW). pp. 90–96. IEEE, Lisbon, Portugal (Sep 2017)
23. Norman, D.: Emotional Design: Why We Love (or Hate) Everyday Things. Basic Books (2004)
24. O'Hare, D.: The 'Wheel of Misfortune': A Taxonomic Approach to Human Factors in Accident Investigation and Analysis in Aviation and Other Complex Systems, vol. 43 (2001)
25. Pereira, D., Hirata, C., Pagliares, R., Nadjm-Tehrani, S.: Towards Combined Safety and Security Constraints Analysis. In: Tonetta, S., Schoitsch, E., Bitsch, F. (eds.) Computer Safety, Reliability, and Security, vol. 10489, pp. 70–80. Springer International Publishing, Cham (2017)
26. Pereira, D.P., Hirata, C., Nadjm-Tehrani, S.: A STAMP-based ontology approach to support safety and security analyses. Journal of Information Security and Applications **47**, 302–319 (Aug 2019)
27. Piètre-Cambacédès, L., Bouissou, M.: Cross-fertilization between safety and security engineering. Reliability Engineering & System Safety **110**, 110–126 (Feb 2013)

28. RAIB: Loss of safety critical signalling data on the Cambrian Coast line. https://www.gov.uk/raib-reports/report-17-2019-loss-of-safety-critical-signalling-data-on-the-cambrian-coast-line (Dec 2019)
29. Reason, J.: Human Error by James Reason (Oct 1990)
30. Schneier, B.: Academic: Attack Trees - Schneier on Security by Dr. Dobb's Journal. https://www.schneier.com/academic/archives/1999/12/attack_trees.html (Dec 1999)
31. Schneier, B.: Secrets and Lies: Digital Security in a Networked World. John Wiley & Sons (2000)
32. Shorrock, S.T.: Errors of perception in air traffic control. Safety Science **45**(8), 890–904 (2007)
33. Shostack, A.: Threat Modeling: Designing for Security. John Wiley and Sons, Indianapolis, IN (2014), adam Shostack
34. Slominski, H.M.: Using STPA and CAST to Design for Serviceability and Diagnostics p. 106 (May 2020)
35. 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)
36. Winther, R., Johnsen, O.A., Gran, B.A.: Security Assessments of Safety Critical Systems Using HAZOPs. In: SAFECOMP (2001)
37. 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)