

**Integrating Safety, Security and Human Factors Engineering in Rail
Infrastructure Design and Evaluation**



By

Amna Altaf

A thesis submitted to the Faculty of Science and Technology, Bournemouth University, UK in partial fulfilment of the requirements for the degree of *Doctor of Philosophy*.

March 2022

Declaration

Confidentiality

I confirm that this dissertation does not contain information of a commercial or confidential nature or include personal information other than that which would normally be in the public domain unless the relevant permissions have been obtained.

In particular any information which identifies a particular individual's religious or political beliefs, information relating to their health, ethnicity, criminal history or sex life has been anonymised unless permission has been granted for its publication from the person to whom it relates.

Original Work Declaration

This dissertation and the project that it is based on are my own work, except where stated, in accordance with University regulations.

Signed:

Amna Altaf
March 2022

Acknowledgements

This PhD research was funded by Bournemouth University (BU) and Ricardo Rail Studentship. I am very grateful to Ricardo Rail for the sponsorship of this work. I am also grateful to the Department of Computing and Informatics at BU for additional sponsorship of fieldwork and conference related activities.

I would like to express my gratitude towards my academic supervisors Dr. Shamal Faily, Dr. Huseyin Dogan and Dr. Alexios Mylonas for their continuous invaluable guidance throughout this PhD. It has been an absolute amazing experience to be able to work under their supervision.

I would like to thank my industrial supervisor Dr. Eylem Thron as well, for her immense support and feedback during this research.

I would like to thank all the staff members at BU, for creating such a friendly and excellent research environment. Especially, I would like to thank Doctoral College and Sci-Tech Staff at BU for their assistance towards research students.

I dedicate this PhD to my beloved parents, my younger brother Omair and sister Anam, who always stood by my side. A special thanks for their love, encouragement, inspiration and support, for all the achievements in my life.

Abstract

With the new emerging dependency towards the rail industry, there have been growing concerns on how to make this critical infrastructure more adaptable in this technological era of cyber attacks. Currently, the rail infrastructure is built around safety and human factors, but one important factor which has less attention is cyber security. In order to satisfy the security needs of rail stakeholders, there is a need to put together knowledge in the form of design framework by combining safety and human factors, with cyber security. The research problem this PhD thesis addresses is how the process-techniques and tool-support available in safety, security and human factors engineering can be integrated to provide design solutions in rail infrastructure.

This PhD thesis claims that proposed design framework is an exemplar by making three significant contributions. Firstly, it identifies the integration of concepts between safety, security and human factors engineering. Secondly, based on integration it provides an integrated design framework where Integrating Requirements and Information Security (IRIS), use-case specifications informed Task Analysis (TA) using Cognitive Task Analysis (CTA) and Hierarchical Task Analysis (HTA), Human Factors Analysis and Classification System (HFACS) frameworks are used to inform Systems-Theoretic Process Analysis (STPA). This integrated design framework is tool-supported using the open-source Computer Aided Integrating Requirements and Information Security (CAIRIS) platform. Thirdly, the proposed design framework in the form of process-techniques and tool-support is implemented by rail infrastructure to determine the safe, secure and usable design solutions.

This PhD thesis is validated by applying the design framework to three case studies. In the first, preliminary evaluation is carried out by applying it to a case study of 'Polish Tram Incident', where inter-dependencies between safety, security, and human factors engineering are present. In the second, the results are used to inform TA using use-case specifications format by prototyping the role of European Railway Traffic Management System (ERTMS) - Signaller, which provides human factors experts a chance to work in collaboration with safety and security design experts. In the final case study, with the support of representative rail stakeholders from Ricardo Rail is used to implement STPA on case study of 'Cambrian Railway Incident'.

Contents

- 1 Introduction 1**
 - 1.1 Research Motivation 1
 - 1.2 Research Objectives 3
 - 1.3 Research Goal 4
 - 1.4 Research Question 4
 - 1.5 Thesis Overview 5
 - 1.6 Thesis Structure 5
 - 1.7 Research Contributions 8
 - 1.7.1 Security Analysis for Safety and Human Factors Issues 8
 - 1.7.2 Use-case Informed TA Using CTA and HTA 9
 - 1.7.3 STPA Alignment with IRIS and CAIRIS 9
 - 1.8 Related Publications 10

- 2 Related Work 11**
 - 2.1 Designing Safety-Critical Systems 11
 - 2.1.1 Hazard Analysis 13
 - 2.1.2 Risk Analysis 13
 - 2.1.3 Fault Tolerance 13
 - 2.1.4 Reliability 13
 - 2.1.5 Verification and Validation 14
 - 2.2 Intersection between Safety & Security Engineering 14
 - 2.2.1 Common Approaches 16
 - 2.2.2 Systems-Theoretic Process Analysis 16
 - 2.2.3 Systematic Review - A STAMP Model 16
 - 2.2.4 Applying STPA in Case Studies 17
 - 2.2.5 STPA by Safety and Human Factors Experts 19
 - 2.3 Security-by-Design & Implementation 20
 - 2.3.1 Basics of Threat Modelling 21
 - 2.3.2 MITRE ATT&CK Framework 22
 - 2.3.3 Potential Human Error 22

| | | |
|----------|---|-----------|
| 2.3.4 | Generic Error Modelling System | 22 |
| 2.3.5 | Human Performance Evaluation | 23 |
| 2.3.6 | Design Techniques | 23 |
| 2.4 | Security leading to Human Factors Engineering | 25 |
| 2.4.1 | Toulmin's Model of Argumentation | 25 |
| 2.4.2 | Use-Case Scenario and Template | 25 |
| 2.4.3 | KAOS - Goal Modelling Language | 26 |
| 2.4.4 | Integrating Requirements and Information Security | 26 |
| 2.5 | Human Factors Design Solutions | 27 |
| 2.5.1 | User-Centered Design Approach | 28 |
| 2.5.2 | Fundamental Design Principles | 28 |
| 2.5.3 | Task Analysis Processes and Tools | 29 |
| 2.6 | Overlap between Human Factors & Safety Engineering | 30 |
| 2.6.1 | Swiss Cheese Model of Accident Causation | 31 |
| 2.6.2 | Human Factors Analysis and Classification System | 32 |
| 2.7 | Rail Infrastructure Design and Evaluation | 33 |
| 2.7.1 | Human Performance and Reliability | 34 |
| 2.7.2 | Tools and Resources | 34 |
| 2.7.3 | Brief Comparison of Standards and Practices in Rail | 36 |
| 2.8 | Summary | 37 |
| 3 | Methodology | 39 |
| 3.1 | Research Approaches | 39 |
| 3.1.1 | Establishing Requirements | 41 |
| 3.1.2 | Design and Prototyping | 41 |
| 3.1.3 | Theoretical and Experimental Approaches | 42 |
| 3.2 | Proposed Research | 42 |
| 3.3 | Application of Research Methods | 42 |
| 3.3.1 | Literature and Systematic Review | 43 |
| 3.3.2 | Theoretical Cyber-Model Framework | 44 |
| 3.3.3 | Interviews | 45 |
| 3.3.4 | Case Study Research | 45 |
| 3.4 | Summary | 47 |
| 4 | Safe, Secure and Usable Design Framework | 48 |
| 4.1 | Integration of Concepts | 48 |
| 4.2 | Security-by-Design Approaches | 50 |
| 4.2.1 | IRIS and CAIRIS | 52 |
| 4.2.2 | Asset Modelling and their Associations | 53 |
| 4.2.3 | Role and Attacker Personas | 53 |

| | | |
|----------|---|-----------|
| 4.2.4 | Vulnerability Identification and Threat Modelling | 54 |
| 4.2.5 | Risk Analysis | 54 |
| 4.2.6 | Task and Goal-Obstacle Modelling | 55 |
| 4.3 | Identification of Safety Hazard | 56 |
| 4.4 | HFACS Framework | 56 |
| 4.5 | Case Study - Polish Tram incident | 57 |
| 4.5.1 | Overview | 57 |
| 4.5.2 | Attacker Perspective | 58 |
| 4.5.3 | IR Remote Control | 59 |
| 4.5.4 | Cyber Attack | 60 |
| 4.6 | IRIS and CAIRIS | 61 |
| 4.6.1 | Asset Modelling and their Associations | 61 |
| 4.6.2 | Role and Attacker Personas | 62 |
| 4.6.3 | Vulnerability Identification and Threat Modelling | 64 |
| 4.6.4 | Risk Analysis | 64 |
| 4.6.5 | Task and Goal-Obstacle Modelling | 65 |
| 4.7 | Identification of Safety Hazard | 66 |
| 4.8 | Human Error - HFACS | 67 |
| 4.9 | Discussion | 67 |
| 4.10 | Summary | 69 |
| 5 | Extension of Design Framework | 71 |
| 5.1 | Human Factors Engineering Techniques | 71 |
| 5.1.1 | Personas for Task Elicitation | 72 |
| 5.1.2 | Use-Case Specifications Informed Task Analysis | 73 |
| 5.1.3 | Cognitive Task Analysis | 74 |
| 5.1.4 | Hierarchical Task Analysis | 75 |
| 5.2 | Implementation in CAIRIS | 76 |
| 5.2.1 | Development Environment | 77 |
| 5.2.2 | Database Tables and Procedures | 78 |
| 5.2.3 | Python Scripting and Graphviz Models | 78 |
| 5.3 | Case Study - ERTMS Signaller | 78 |
| 5.3.1 | Overview | 78 |
| 5.3.2 | Task Breakdown | 79 |
| 5.3.3 | Use-case View | 80 |
| 5.4 | Task Analysis | 81 |
| 5.4.1 | Personas for Task Elicitation | 81 |
| 5.4.2 | Use-Case Specifications Informed Task Analysis | 82 |
| 5.4.3 | Cognitive Task Analysis | 83 |

| | | |
|----------|---|------------|
| 5.4.4 | Hierarchical Task Analysis | 83 |
| 5.4.5 | Risk Analysis | 84 |
| 5.5 | Discussion | 84 |
| 5.6 | Summary | 85 |
| 6 | Integrated Design Framework for Facilitating STPA | 87 |
| 6.1 | Safety Analysis | 87 |
| 6.2 | STPA Process Model | 88 |
| 6.2.1 | Pre-requisite | 88 |
| 6.2.2 | Step 1: Accident, Hazard and Constraint | 89 |
| 6.2.3 | Step 2: Model Control Structure | 90 |
| 6.2.4 | Step 3: Unsafe Control Action | 90 |
| 6.2.5 | Step 4: Causal Factor | 90 |
| 6.2.6 | Step 5: Risk Analysis Model | 91 |
| 6.3 | Case Study - Cambrian Railway Incident | 91 |
| 6.3.1 | Overview | 91 |
| 6.3.2 | Breakdown of Events | 92 |
| 6.3.3 | Choice of Incident | 93 |
| 6.4 | Partial-STPA Assessment | 93 |
| 6.4.1 | Pre-requisite | 96 |
| 6.4.2 | Step 1: Accident, Hazard and Constraint | 99 |
| 6.4.3 | Step 2: Model Control Structure | 100 |
| 6.4.4 | Step 3: Unsafe Control Action | 101 |
| 6.4.5 | Step 4: Causal Factor | 101 |
| 6.4.6 | Step 5: Risk Analysis Model | 101 |
| 6.5 | Discussion | 102 |
| 6.6 | Meta-Model of Design Framework | 103 |
| 6.7 | Summary | 105 |
| 7 | Conclusion | 107 |
| 7.1 | Evaluation of Research Questions | 107 |
| 7.1.1 | RQ1 - Integration of Concepts | 107 |
| 7.1.2 | RQ2 - Process-Techniques & Tool-Support | 108 |
| 7.1.3 | RQ3 - Design Framework | 109 |
| 7.2 | Key Research Findings | 110 |
| 7.2.1 | Bridging Safety, Security and Human Factors | 111 |
| 7.2.2 | Implementation of Design Framework | 111 |
| 7.3 | Research Challenges and Limitations | 113 |
| 7.4 | Future Work | 114 |
| 7.4.1 | Application of Safe, Secure and Usable Design Framework | 114 |

| | |
|--|------------|
| 7.4.2 Industrial Viewpoint | 114 |
| References | 115 |
| Appendices | 123 |
| A Ethics Approval | 124 |
| B Interview Process & Template | 125 |
| C Case Studies - CAIRIS Model Files | 127 |

List of Figures

| | | |
|------|---|----|
| 1.1 | Thesis Overview | 6 |
| 2.1 | Hazard Analysis Cycle (HA 2019) | 12 |
| 2.2 | Verification and Validation Testing (Storey 1996) | 14 |
| 2.3 | Safety and Security Engineering along with Human Factors | 15 |
| 2.4 | STPA Model from STAMP (Karatzas and Chassiakos 2020) | 17 |
| 2.5 | STPA Method Derivation from Control Flaw Process (Karanikas 2016) | 19 |
| 2.6 | Task Analysis Approach Association with STPA Method (Karanikas 2016) | 20 |
| 2.7 | IRIS Framework Meta-Model Views (Faily 2018) | 27 |
| 2.8 | Swiss Cheese Model of Accident Causation (Reason 1990) | 31 |
| 2.9 | HFACS Framework (Zhou and Lei 2018) | 32 |
| 2.10 | Cognitive Reactions Responsible for Human Performance (Hammerl and Vanderhaegen 2009) | 33 |
| 2.11 | Security, Safety and Human Factors Existing Approaches | 37 |
| 3.1 | Qualitative Research Design Cycle (Steeves 2018) | 39 |
| 3.2 | Action Research and Grounded Theory Cycle (Chen and Cheng 2015) | 41 |
| 3.3 | Application of Research Methods for Thesis | 43 |
| 3.4 | Literature and Systematic Review Process | 44 |
| 3.5 | Qualitative Case Study Research Plan (Harrison et al. 2017) | 46 |
| 4.1 | Integration of Concepts Between Safety, Security and Human Factors | 49 |
| 4.2 | Security by Design Approach Consisting of IRIS Framework | 50 |
| 4.3 | CAIRIS Graphical Notation | 52 |
| 4.4 | Asset Model Using UML | 53 |
| 4.5 | Argumentation Model for Personas Characteristic | 54 |
| 4.6 | Threat and Risk Analysis Modelling in CAIRIS | 55 |
| 4.7 | Task and Goal-Obstacle Modelling in CAIRIS | 55 |
| 4.8 | High-level Architectural Overview for Polish Tram System | 58 |
| 4.9 | Working Infrastructure of Polish Tram System | 58 |
| 4.10 | Hardware Setup for IR Remote Control (Ard 2009) | 59 |

| | |
|--|-----|
| 4.11 Cyber Attack for Polish Tram System | 60 |
| 4.12 Asset Model Using UML | 61 |
| 4.13 Argumentation Model for Personas Characteristic | 64 |
| 4.14 Risk Modelling in CAIRIS | 65 |
| 4.15 Task and Goal-Obstacle Modelling in CAIRIS | 66 |
| 5.1 UML for Use-Case Specifications Informed Task Analysis | 72 |
| 5.2 Use-case Specification Template for Task Analysis | 73 |
| 5.3 CAIRIS Architecture (Faily 2018) | 77 |
| 5.4 Use-Case View for ERTMS Signaller | 80 |
| 5.5 Use-Case Specification for 'Conflict Prediction and Resolution' | 82 |
| 5.6 HTA Graph with Levels of Human Failure | 84 |
| 6.1 Accident, Hazard and Constraint Model using Knowledge Acquisition in autOmed Specification (KAOS) Association in CAIRIS | 89 |
| 6.2 Model Control Structure using Data Flow Diagram (DFD) in CAIRIS | 90 |
| 6.3 Route for Cambrian Coast Line (Les 2019) | 92 |
| 6.4 Activity Diagram for STPA Using Design Framework | 94 |
| 6.5 Goal-Obstacle Model for Cambrian Incident Case Study | 95 |
| 6.6 Persona Characteristic of 'Attitudes' for Neil in CAIRIS | 96 |
| 6.7 Task Participation Form for 'Self-Test Function' in CAIRIS | 97 |
| 6.8 Actor Identification for 'Operational Planning' Use-case in CAIRIS | 98 |
| 6.9 KAOS Association Between Accident and Hazard | 99 |
| 6.10 KAOS Association Between Accident, Hazard and Constraint | 99 |
| 6.11 High-level Control Structure Model | 100 |
| 6.12 DFD of Control Structure Model using CAIRIS | 100 |
| 6.13 Risk Analysis Model Based on Attacker, Threat and Vulnerability | 102 |
| 6.14 Meta-Model of Design Framework | 104 |
| B.1 Template of Task Sheet for Interview | 125 |
| C.1 GitHub repository for CAIRIS Models | 127 |

List of Tables

| | | |
|-----|--|-----|
| 2.1 | Methods and Tools for Task Analysis with Applications | 29 |
| 2.2 | Brief Comparison of Standards and Practices in Rail | 36 |
| 3.1 | Case Study Research for PhD | 46 |
| 4.1 | Online Articles Used as Data Source for Building Attacker Personas | 62 |
| 4.2 | Major Roles in Rail Infrastructure | 63 |
| 4.3 | Human Factors Issues based on HFACS | 67 |
| 5.1 | Cognitive Reactions and Performance Shaping Factors | 73 |
| 5.2 | Documentation and Literature used for Train Signaller Personas | 81 |
| 5.3 | Cognitive Task Analysis for Use-Case Specifications | 83 |
| 6.1 | Literature Survey on Cambrian Incident | 93 |
| 6.2 | Unsafe Control Action corresponding to Accident, Hazard and Constraint . | 101 |
| A.1 | Research Ethics Checklist | 124 |

Acronyms

ACSE Assurance and Safety Case Environment.

ALARP As Low As Reasonably Practicable.

ATCPS Advanced Train Control Protocol System.

ATO Automatic Train Operation.

BU Bournemouth University.

CAE Claims, Arguments and Evidence.

CAIRIS Computer Aided Integrating Requirements and Information Security.

CAPEC Common Attack Pattern Enumeration and Classification.

CASE Computer Aided Software Engineering.

CoPs Codes of Practice.

CSM-REA Common Safety Method for Risk Evaluation and Assessment.

CTA Cognitive Task Analysis.

CTT Concur Task Trees.

DFD Data Flow Diagram.

DiD Defence-in-Depth.

DMI Driver Machine Interface.

ERTMS European Railway Traffic Management System.

ETCS European Train Control System.

EU European Union.

FMEA Failure Mode and Effects Analysis.

- FOQA** Flight Operation Quality Assurance.
- GEMS** Generic Error Modelling System.
- GSM-R** Global System for Mobile communications for Railways.
- GSN** Goal Structuring Notation.
- GUI** Graphical User Interface.
- HAZOP** Hazard and Operability Study.
- HCI** Human Computer Interaction.
- HCI-security** Human Computer Interaction - Security.
- HERMES** Human Error Risk Management for Engineering Systems.
- HFACS** Human Factors Analysis and Classification System.
- HFRM** Human Factors Risk Manager.
- HFW** Human Factors Workbench.
- HMI** Human Machine Interface.
- HTA** Hierarchical Task Analysis.
- IEC** International Electro-technical Commission.
- IMEA** Intrusion Modes and Effects Analysis.
- IoT** Internet of Things.
- IR** Infra-red.
- IRIS** Integrating Requirements and Information Security.
- ISAs** Independent Safety Assessments.
- ISO** International Organization for Standardisation.
- IT** Information Technology.
- KAOS** Knowledge Acquisition in autOmated Specification.
- LED** Light-emitting Diode.
- NCSC** National Cyber Security Center.

ORR Office of Rail Regulation.

OT Operational Technology.

OWASP Open Web Application Security Project.

PHEA Predictive Human Error Analysis.

PIFs Performance Influencing Factors.

PSFs Performance Shaping Factors.

RAIB Rail Accident Investigation Branch.

RAMS Reliability, Availability, Maintainability and Safety.

RBC Radio Block Center.

RSSB Railway Safety and Standards Board.

STAMP Systems-Theoretic Accident Model and Processes.

STPA Systems-Theoretic Process Analysis.

STPA-Sec STPA for Security.

TA Task Analysis.

TNA Training Needs Analysis.

TSR Temporary Speed Restriction.

UIC International Union of Railways.

UK United Kingdom.

UML Unified Modelling Language.

UX User Experience.

XML Extensible Markup Language.

Chapter 1

Introduction

In this chapter, the motivation behind research problem is stated. This is followed by research objective and purpose for this PhD. The research goal is accomplished by answering three research questions. An overview of thesis structure has been given, along with research contributions and a list of publications made during this PhD.

1.1 Research Motivation

Protecting the health and safety of passengers is critical in the rail industry, which is evolving to meet new passenger and freight demands. The evolution is focused around the automation of railway processes like the introduction of Automatic Train Operation (ATO) for operational safety enhancement, European Railway Traffic Management System (ERTMS) for signalling and speed control, European Train Control System (ETCS) for automatic train protection, and Global System for Mobile communications for Railways (GSM-R) as operating standard.

Traditionally, the rail infrastructure is built around safety and human factors. However, as the rail information infrastructure becomes integrated with Operational Technology (OT), especially with the implementation of ERTMS, new vulnerabilities are introduced together with the new threats that exploit them. These vulnerabilities are as a result of the dependence of OT on a network support infrastructure, which provides various Internet gateways as opportunities for attackers. The attackers scrutinise these opportunities by looking for hidden weaknesses leading to attack scenarios such as a threat or risk. As such attacks are directly or indirectly responsible for compromising safety, cyber security as well has become a new concern for rail safety engineers.

The European Union Agency for Cyber Security (ENISA) aims to fulfil cyber security goals within rail infrastructure by nominating cyber security as an essential requirement for rail sector (European Network and Information Security Agency 2020). The following are the major characteristics and challenges of the rail industry domain, which need to

be recognised when dealing with cyber security:

Long Life-cycle of Products: The technologies behind rail products are updated over decades. Usually, a rail upgrade plan is devised by keeping in mind a multi-billion budget. In this situation, recommending a swift cyber security design change is challenging and more demanding in terms of infrastructure requirements, delivery of service, time and budget scaling, and other management regulations etc.

Legacy Systems: The operational efficiency behind legacy systems within rail can be improved by providing digital solutions. These digital solutions are focused around safety, where security is an emerging concept. Therefore, while planning digital transformation of legacy systems the cyber security needs to be included as well.

Strong Safety Culture & Lack of Security: The foundation of rail industry is built around safety. The European Railway Safety Culture Model states the design and implementation strategies based on cultural enablers, behaviour patterns and railway safety fundamentals (Ouferroukh 2018). There is a need to enable a strong security culture adjacent to safety for rail industry.

Cyber Security Awareness for Staff: For example, the incident of WannaCry virus within Deutsche Bahn systems, where ransom money was demanded. The key consideration should be given to the fact that cyber response team was able to catch virus but they lacked apt response and recovery plans (Tech 2017).

Overlap Between Safety and Security: Efforts should be made by stakeholders to reconcile the gaps between safety and security. However, there are challenges where the safety-critical systems being implemented by safety engineers need to be revised keeping in mind the security aspects.

Involvement of Internet of Things (IoT) within Rail: As a result of this digital transformation, there are chances for threats due to new and unknown vulnerabilities. The potential attackers may compromise system safety and security by exploiting these exposed vulnerabilities.

Service Efficiency with Security: Implementation of cyber security measures, by keeping in mind the cost-benefit analysis. Majority of rail infrastructure is built for public, where providing service with affordability is the main goal.

Usually the railways are guarded by Independent Safety Assessments (ISAs), whose general concern is coming to a decision about the safety of the system. However, such decisions also depend on the security assessment made against the expected attacks. The ISAs are conducted to determine the possible causes behind the hazards and accidents, which include security breaches. A *hazard* is defined as an act or set of activities

which has the potential (risk) to cause an injury or damage whereas an *accident* is unknown event (occurrence) leading to unfortunate circumstances (Storey 1996).

Often poor design decisions made during security engineering may lead operators to make human errors or mistakes where rules are un-intentionally disobeyed (Reason 1990). This may eventually affect system safety. Thus, system safety is compromised due to human intervention in the form of errors and mistakes. The human error is considered as the biggest source for active failures. Normally, active failures lead to accidents and incidents, whereas the latent failures usually lie beneath these active failures and may lead to same catastrophic events causing harm to human life.

Rail infrastructure should be strong enough to block opportunities of human error, but not at the cost of security and safety. For example, work by (Cacciabue 2005) describes how the Human Error Risk Management for Engineering Systems (HERMES) risk management approach assesses the chance of human error, but while it identifies safety and reliability components required in railways, it does not consider cyber security. Similarly, the Generic Risk Assessment Log presented by Randstad Rail mentions all the possible events that can lead to safety hazards, but does not mention the associated security concerns (Ran 2014).

Integrating Requirements and Information Security (IRIS) framework and Computer Aided Integrating Requirements and Information Security (CAIRIS) tool-support have been used in several real-world case studies, including the development of security policies for critical infrastructure systems (Faily and Flechais 2011). The IRIS framework was devised to understand how design concepts associated with security, usability, and software engineering could be aligned (Faily 2018). Using IRIS, multiple views of system in the form of environment, asset, task, goal, and risk can be put forward for analysis. This gives both security (risk analysis) and human factors experts (usability and cognitive attributes) to contribute together. However, the safety inclusion remains a riddle to be solved.

Therefore, rail infrastructures can only be made strong if along with safety and human factors, the security engineers contribute to its design and evaluation. To do this, we need to formulate a design framework based on approaches that integrate security engineering with safety and human factors engineering. The problem this research work addresses is how the process-techniques and tool-support available in safety, security and human factors engineering can be integrated in the form of design framework to provide solutions for rail infrastructure.

1.2 Research Objectives

The following are the major research objectives:

1. Integrate security along with safety and human factors in rail. For this purpose, the concepts (variables) to be considered as foundation of integration are: safety hazard, security risk, and human error. These concepts will be used to bridge the gap between three domains.
2. Define an integrated design framework based on process-techniques and tool-support from safety, security and human factors engineering. This design framework will aim to provide safe, secure and usable design solutions.
3. Conduct security risk analysis leading to potential safety hazards and human factors issues. This will ensure the applicability of the design framework within the rail infrastructure.

These research objectives will be achieved by identifying the research goal and research questions, while keeping in mind the research methodologies and possible limitations.

1.3 Research Goal

The aim of this PhD research is to come forward with a design framework in the form of process-techniques and tool-support based on an integration of concepts from safety, security and human factors engineering to provide design solutions in rail infrastructure. The idea is to provide a single platform in the form of design framework where safety, security and human factors experts can insert their individual inputs and as a result, are able to look at integrated design solutions.

1.4 Research Question

To achieve this multi-disciplinary critical infrastructure research goal, three major research questions are answered:

RQ1 - Integration of Concepts: How the concepts from safety, security and human factors engineering can be integrated together to build the foundation for design framework?

RQ2 - Process-Techniques & Tool-Support: How the processes and techniques based on safety, security and human factors engineering integration can be adopted, along with available tool-support options to propose a new design framework?

RQ3 - Design Framework: How by the application of proposed design framework the safety, security and human factors engineering design concerns are resolved in rail?

1.5 Thesis Overview

This PhD thesis is used to address integration of safety, security and human factors engineering for resolving design issues in rail infrastructure. The thesis is broken down into the subsequent seven chapters as shown in Fig. 1.1. Chapter 2 reviews all available process-techniques from safety, security and human factors engineering, along with tool trends available in rail. Chapter 3 states the research methodology applied during this research. Chapter 4 presents the integrated design framework, which is validated by the case study. Chapter 5 briefs about the extension of design framework along with case study application. Chapter 6 facilitates the STPA process-model for a case study using integrated design framework. Finally, in Chapter 7, the thesis is evaluated and concluded.

1.6 Thesis Structure

The major aim and goal of this PhD research project as motivated by the research problem in Section 1.1, is accomplished by answering three relevant research questions as stated in Section 1.4. In rail industry, sometimes security incidents could potentially have safety implication. Here, human factors especially in the form of human tendency to make errors and mistakes within system has a co-relation with both safety and security. As human error has a certain tendency to compromise the security of system and eventually safety of people within an environment. Safety, security and human factors engineering techniques are largely disconnected, although there is a certain overlap between the concepts. Based on that overlap, we need process-techniques that integrate security engineering with safety and human factors engineering.

In Chapter 2, relevant literature is reviewed towards understanding the basics of safety-critical systems and their design factors, security-by-design techniques, and human factors approaches. Here, a co-relation between safety and security, security and human factors, and human factors and safety has been determined based on available research. The rail infrastructure codes of practice and standards used to implement safety and human factors has also been mentioned, along with available tool-support options. This chapter aims to address RQ1 where the integration between safety, security and human factors engineering is needed. Based on this identified integration, the choice of process-techniques and tool-support is made for the design framework. Thus, this chapter also addresses RQ2 and in part lays foundation for RQ3.

In Chapter 3, a comprehensive overview of available research approaches is given. Based on this, the choice of methodology for conducting this proposed research work has been made.

In Chapter 4, the safe, secure and usable design framework is presented. The core

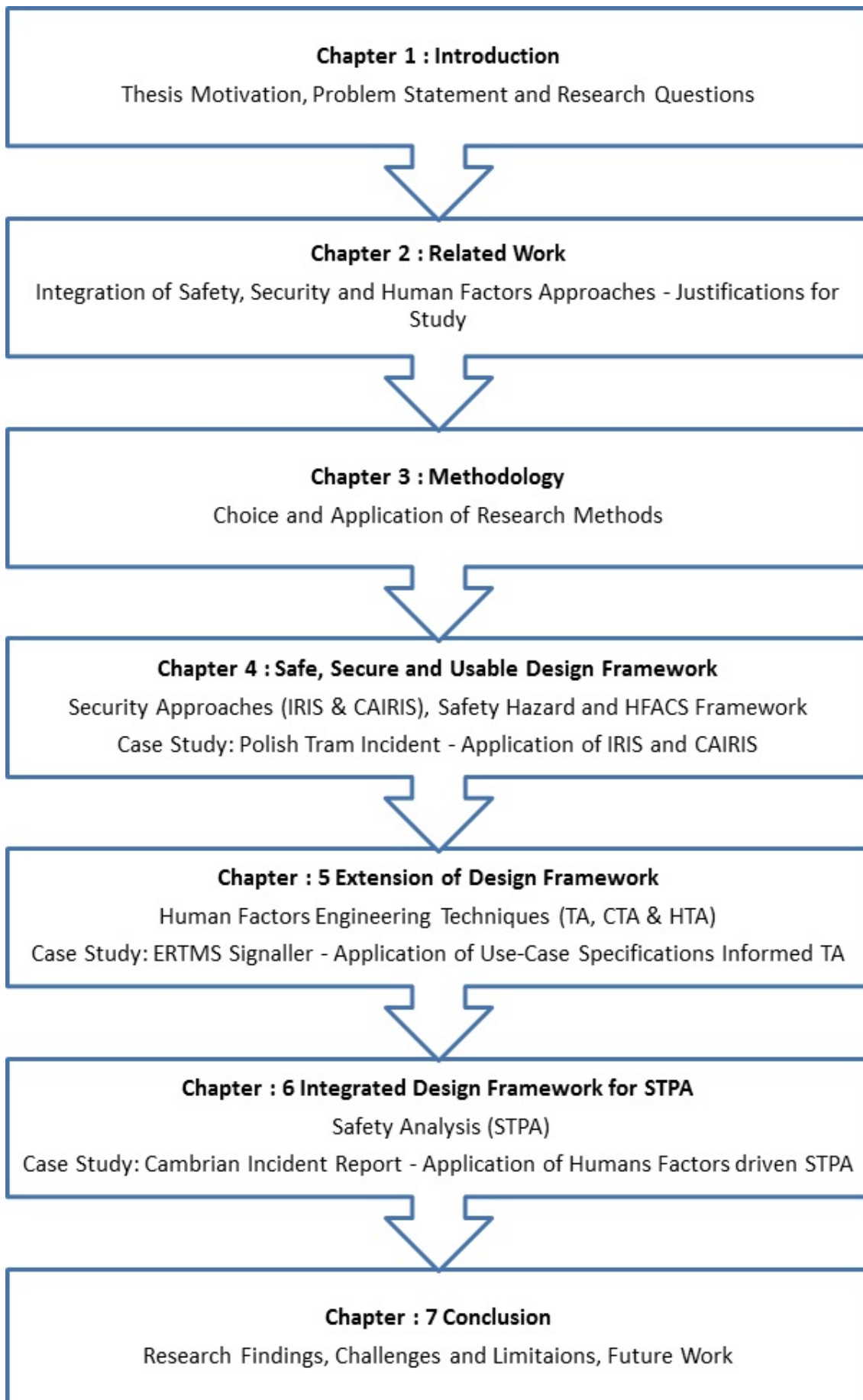


Figure 1.1: Thesis Overview
Faculty of Science & Technology, Bournemouth University - PhD

concepts from the IRIS framework are used to define an intersecting model, based on a proposed relationship between different security-by-design and usability techniques. The framework is tool-supported using the open-source CAIRIS platform. Also, the HFACS framework has been used in conjunction with IRIS and CAIRIS, to determine the appropriate human error sources and potential safety hazards. An evaluation of design framework is conducted by applying it to a real-life case study of 'Polish Tram Incident', where inter-dependencies between safety, security, and human factors engineering were present. In doing so, three significant contributions are made. First, the proposed process-technique 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, it has 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) were determined using HFACS framework. This chapter aims to address RQ2 and RQ3. For the evaluation and improvement of this proposed process-technique the representative rail stakeholders from Ricardo Rail were closely involved when considering the risks, roles, tasks, goals, requirements, dependencies and obstacles between the humans and systems. Based on their experiences and feedback reviews, improvements to proposed process-technique were done. This chapter aims to address RQ1 and RQ2, where the available process-techniques and tool-support options are adopted into design framework based on an integration scope between safety, security and human factors engineering.

In Chapter 5, the design framework is extended. The humans factors engineering technique such as TA is conducted using a use-case specification pre-defined format in CAIRIS. For each use-case specification CTA and HTA is performed. CTA is conducted by scoring relevant cognitive reactions. This leads to identification of different levels of human failures with the use of *Algorithm 1*. During HTA, associations between use-cases are identified. After colour coding of the use-cases, graphical models are generated using *Algorithm 2*. Moreover, the graphical models generated by CAIRIS are used to determine potential safety hazards. These safety hazards are then used to conduct STPA, for identifying control actions and causal factors behind accidents for improving system design. For demonstration of use-case specifications informed TA, the role (i.e. Signaller) using ERTMS is prototyped. A preliminary evaluation is done of regular tasks performed by an 'ERTMS Signaller', which highlights human error sources behind these tasks. In doing so, three contributions are made. First, TA approach is derived from security and requirements engineering IRIS framework using concepts such as roles and personas, task and goal-obstacle modelling. Second, TA is applied as a combination of CTA and HTA tool, highlighting the importance of mental load with a detailed task breakdown. Finally, these

tools are applied using use-case specifications template, thus providing task sequence with exception identification. These exceptions help security and safety experts to conduct risk and hazard analysis, by identifying potential vulnerabilities and threats hidden beneath system design. This chapter aims to address RQ2 and RQ3.

In Chapter 6, STPA method is applied for a case study of 'Cambrian Incident' by using human factors and security approaches from integrated design framework. The human factors approach such as identification of roles and personas, task modelling and use-cases are used to understand processes, asset associations and goal-obstacle models. In return, goal-obstacle models and DFD (processes and datastores) are used to conduct STPA, where risk analysis based on recognition of attacker/s, threats, vulnerabilities, risks and misuse cases are done, simultaneously. All these process-techniques are tool-supported by open-source CAIRIS platform. This helps to understand an integration of concepts between safety and security, security and human factors, and human factors and safety. Thus, laying a foundation of an overlap of concepts between three domains, which leads to recognition of safe, secure and usable design framework. Using this integrated design framework, safety goals, security risks and human factors concerns are highlighted. Also, by tool-support the effort required by safety, security and human factors experts is minimised by providing automated and efficient design solutions. This chapter aims to address RQ3.

In Chapter 7, a discussion about general findings and observations about the conducted research work has been made with particular interest towards cyber security, potential safety hazards and human factors issues. A conclusive summary has been given about evaluation of research questions. The research challenges and limitations faced during the course of this PhD research have also been highlighted. In future work, a more refined design framework based on process-techniques and tool-support from integrated safety, security and human factors engineering concepts will be presented. For this purpose, an application of design framework in industry is in progress.

1.7 Research Contributions

The research contributions intend to define the scope of integration for recognition of design framework. As a result of this PhD, following knowledge contributions have been made:

1.7.1 Security Analysis for Safety and Human Factors Issues

This PhD has looked on how asset modelling and their associations, can be used to identify security attributes namely, confidentiality, integrity, availability of assets as prioritised by rail stakeholders. This has been shown how building models of attackers contributes

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 framework.

1.7.2 Use-case Informed TA Using CTA and HTA

A TA approach has been derived from the security and requirements engineering IRIS framework using concepts such as roles and personas, task and goal-obstacle modelling. It has been shown how CTA and HTA can be combined as single, tool-support TA approach to highlight the importance of mental load with a detailed task breakdown. Finally, it has shown how use-case specifications assist with task sequencing and exception identification. These exceptions help security and safety experts to conduct risk and hazard analysis by identifying potential vulnerabilities and threats hidden beneath system design.

1.7.3 STPA Alignment with IRIS and CAIRIS

This research work demonstrated how the STPA process model has aligned with IRIS and CAIRIS, providing a single platform for all elements and contributing factors related to hazard analysis. These elements comprised of accident (loss), hazard, system constraint, component (control algorithm), process (mental) model, unsafe control action (obstacle) leading to causal factors. This has shown 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 has enlisted 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 for Security (STPA-Sec). Furthermore, the human error sources has the tendency to contribute towards potential safety hazards. The approach focused on bringing security and human factors methods support to STPA. Initially, the STPA process model was suggested by keeping in mind the safety where several case study applications suggested the involvement of human element. This human element was considerable in a socio-technical environment, where the system weaknesses (vulnerabilities) were highlighted by recognising human error sources. These human error sources have established grounds for understanding potential hazard scenarios and model better risk analysis. Hence, this research has built the scope of connection and integration between safety, security and human factors.

1.8 Related Publications

The research conducted during this PhD has been published as research papers in the conference/ workshop proceedings and presented at doctoral symposiums as stated below:

Conference/ Workshop Proceedings

1. Altaf, A., Faily, S., Dogan, H., Mylonas, A. and Thron, E. Identifying Safety and Human Factors Issues in Rail using IRIS and CAIRIS. In: 5th Workshop On The Security Of Industrial Control Systems and Of Cyber-Physical Systems (CyberICPS 2019) 23-27 September 2019 Luxembourg, Luxembourg. Springer.
2. Altaf, A., Thron, E., Faily, S., Dogan, H. and Mylonas, A. Evaluating the Impact of Cyber Security and Safety with Human Factors in Rail using Attacker Personas. In: Institution of Railway Signal Engineers - Aspect (IRSE 2019) 22-24 October 2019 Delft, The Netherlands.
3. Altaf, A., Faily, S., Dogan, H., Mylonas, A. and Thron, E. Use-Case Informed Task Analysis for Secure and Usable Design Solutions in Rail. In: 16th International Conference on Critical Information Infrastructures Security (CRITIS 2021) 27-29 September 2021 Lausanne, Switzerland. Springer.
4. Altaf, A., Faily, S., Dogan, H., Thron, E. and Mylonas, A. Integrated Design Framework for Facilitating Systems-Theoretic Process Analysis. In: 7th Workshop On The Security Of Industrial Control Systems and Of Cyber-Physical Systems (CyberICPS 2021) 04-08 October 2021 Darmstadt, Germany. Springer.

Doctoral Symposiums

1. Altaf, A. Identifying Safety and Human Factors Issues in Rail using IRIS and CAIRIS. In: London Doctoral Symposium - British Computer Society (BCS 2019) 06 June 2019 London, United Kingdom.
2. Altaf, A. Integrating Safety, Security and Human Factors Engineering in Rail Infrastructure Design and Evaluation In: 11th Annual Postgraduate Research Conference - Doctoral College Bournemouth University 04 December 2019 Bournemouth, United Kingdom.

Chapter 2

Related Work

In this chapter, the concepts from safety, security and human factors engineering are explained to determine an overlap and identify the scope of an integration. The chapter begins by identifying the present design approaches for safety-critical systems. This leads to recognition of common approaches prevalent between safety and security engineering. The basics of threat modelling are analysed for understanding the security-by-design techniques, helping to realise the overlap between security and human factors engineering. In return, this overlap is used to explain the user-centered design approaches between human factors and safety engineering.

The integration between safety and security, security and human factors, and human factors and safety is utilised to build the foundation for the desired safe, secure and usable design framework. Also, the present state-of-the-art for these three engineering domains for rail infrastructure in the form of available process-techniques and tool-support options are reviewed, by highlighting the limitations in existing work. At the end, the chapter is concluded with the co-related design approaches between three domains.

2.1 Designing Safety-Critical Systems

The need for safety-critical systems arises because of everyday technological usage which involves embedded systems and associated applications. Due to this dependence a single minute failure may lead to harm or loss of human life. For example, one could consider the broad spectrum usage of airline systems. A minor mistake may lead to air crash and loss of valuable human lives. Similarly, one minute ignorance in critical-data involved in signalling system within rail, will end up compromising human life. Thus the safety of a system is defined as:

*"The property of a system that when put into use will not endanger human life
(Storey 1996)."*

The design of a safe to use system revolves around some safety properties. These properties are termed as: reliability, availability, integrity, maintainability, dependability, and system recovery. The nature of process chosen is considered critical for production of a safe system. The V-Model of design is usually put into use for safety designing due to its simplicity. The V-Model constitutes of verification and validation for subsequent steps involved in development cycle. But sometimes this model is not adoptable, due to inevitable number of iterations (validations) required during design phase (Storey 1996).

Another example is International Electro-technical Commission (IEC) 1508, where the safety life cycle is performed by considering plan of safety related systems and their risk reduction in parallel. Here, the mechanism for verifying the results of each of the activities relevant to safety is determined. Usually, the safety experts recommend incorporation of safety requirements with user requirements at a stage known as 'Preliminary Hazard Analysis'. The design phase can be divided in to several layers. At the top level decomposition of requirements may occur which eventually leads to development of modules (Brazendale 1995).

The major safety factors which need to be considered for designing safety-critical systems are hazard analysis, risk analysis, fault tolerance, reliability, verification and validation. These are discussed in the following sub-sections as:



Figure 2.1: Hazard Analysis Cycle (HA 2019)

2.1.1 Hazard Analysis

The *hazard* is defined as a property, function or component of a system which has some associated level of risk (Raspotnig and Opdahl 2012). The hazard analysis involves the inspection of consequences that may occur in case of failure of a specific component in a system (Storey 1996). Its task is to determine the probability of occurrence of hazards. For this purpose, the tree analysis of tracking down events and their possible hazards is considered as a good practice. The *Probabilistic Hazard Analysis* and *Failure Mode and Effects Analysis (FMEA)* are two most commonly used practices. Generally, hazard analysis revolves around: hazard identification, hazard listing, hazard critical analysis, hazard log and safety plan recognition (HA 2019) as shown in Fig. 2.1.

2.1.2 Risk Analysis

The product of severity of a failure and the frequency of its occurrence is defined as *risk* in safety-critical systems (Wiley 2013). The risks in broader categories are divided into achieved and tolerable risks. The risks can only be diminished by obtaining certain levels of system integrity (Storey 1996). The integrity levels are like safety ratings which are achieved during the development life cycle of a system. The functional safety standard IEC 61508 defines analysis, realisation and operation as the main stages for achieving safety integrity levels in a system (O’Riordan 2015).

2.1.3 Fault Tolerance

The *fault* in a system is categorised according to its nature, duration and the particular degree of occurrence (Storey 1996). Mostly, the software faults may be due to requirement specification faults, coding faults or logical errors. All fault tolerance methods are based on some form of redundancy. The *Triple Modular Redundancy* for fault removal is the most common method. The functionality checking, consistency checking, instruction monitoring and loop-back testing are some of the best fault detection techniques. Another technique where different results of systems are compared, when provided with same inputs but different versions of same program is known as *N-version Programming*. Here the processing cost required is a point of consideration (Leveson 1987).

2.1.4 Reliability

The number of faults tolerated versus the time, are the two main parameters for *Software Reliability Prediction*. The Combinational and Markos Models are put into use for dealing with reliability issues within a system. The fault tolerance ability of a system is directly responsible for reliability as well (Rausand 2014).

2.1.5 Verification and Validation

The verification and validation goals are achieved through testing. The *verification* is the conformance to the specifications stated at the time of requirements whereas the *validation* is the conformance of specifications according to the customers needs. There are different types of testing: dynamic, static, functional, structured, and random etc. The test-cases are generated for accessing the performance, stress, error-guessing, memory and timing of the system. The control flows and data flows (walkthroughs, check-lists and inspections) are also analysed as formal verification models. The modelling of overall safety validation plan is important. The choice of programming language, supporting tools, expertise and system architecture may contribute towards safety-critical systems (Storey 1996).

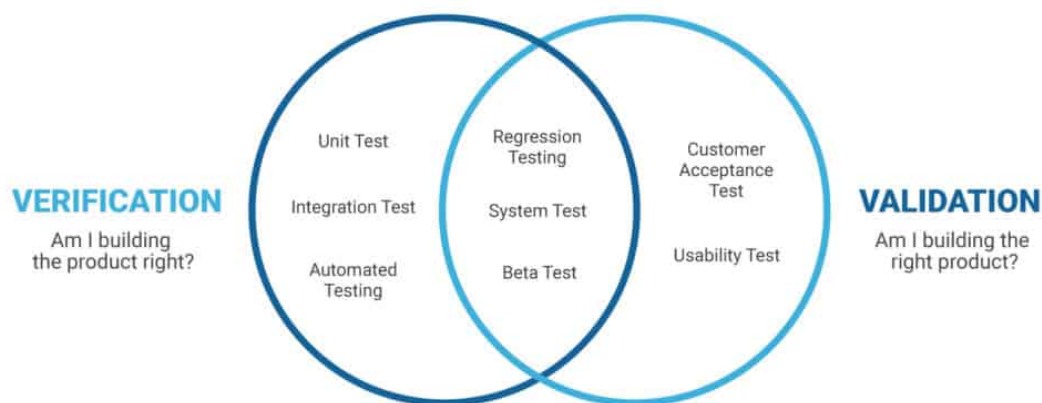


Figure 2.2: Verification and Validation Testing (Storey 1996)

After explaining these safety factors, the focus is shifted towards some terminologies which are highlighted such as hazard and risk. This leads to an identification of the security aspect for system design as minimising risk is prevalent in security as well. Therefore, the safety and security engineering as an overlapping concept are studied in up-coming sections.

2.2 Intersection between Safety & Security Engineering

In general, risk is an intersecting concept between safety and security. Malicious risk is defined as security challenge whereas the accidental risk is defined as safety hazard (Young and Leveson 2014). Malicious risk may have safety implications, such that safety and security can be complementary (Kriaa et al. 2015), and ISAs entail minimising the security threats. At some point the safety and security are considered alike as they both

focus on saving a system while it is in production phase. Hence, somehow safety has similar goals to security.

Hazards and accidents may occur due to security breaches, and dependability – delivering services that can justifiably be trusted – encompasses safety and some major elements of security (Avizienis et al. 2004). Safety is an attribute of dependability, with availability, reliability, integrity and maintainability; security refers to the availability and integrity attributes and to confidentiality (Piètre-Cambacédès and Bouissou 2013). 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.

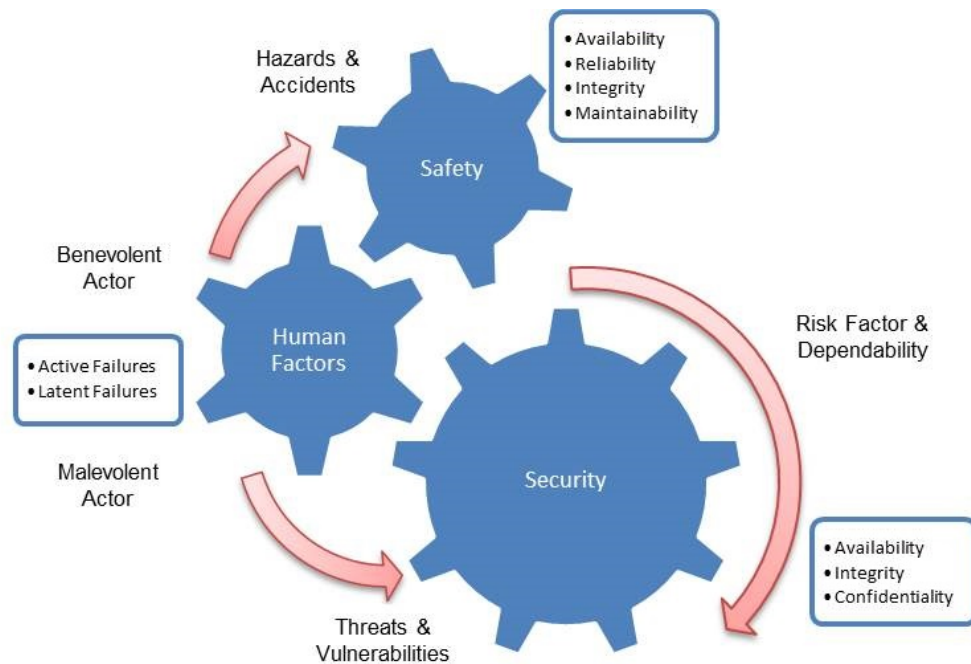


Figure 2.3: Safety and Security Engineering along with Human Factors

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) (Brostoff and Sasse 2001), 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 (Young and Leveson 2014), thereby leading to a risk instigating a safety hazard. This interesting inter-link between safety and security along with human factors engineering concepts as stated above is also shown by the help of Fig. 2.3.

Both safety and security engineering communities are now working to better bridge their communities (Jonsson and Olovsson 1998), e.g. safety engineering consideration of *security mindedness* (Bloomfield et al. 2018). Even the IEC has suggested a framework TC 65/AHG 1 for coordinating safety and security together (IEC 2019).

2.2.1 Common Approaches

There are some existing approaches in safety and security engineering which are common to both due to inter-linked concepts. The Defence-in-Depth (DiD) approach which is meant to delay the attack possibilities by adding additional defence layers and now used to implement security was once derived from a safety design of nuclear plants (Piètre-Cambacédès and Bouissou 2013). In security, the graphical representation of asset attacks related to attackers are shown using attack trees. The concept of attack trees have been derived from fault trees for safety of systems where root-cause of failure is analysed by mapping undesirable events (Schneier 1999). A Hazard and Operability Study (HAZOP) is a structured and systematic approach used to identify and evaluate risk problems in safety. The concept was implemented for security because of risk dealing with security properties (confidentiality, integrity, availability) was discovered as a linking factor (Winther et al. 2001). Similarly, FMEA approach from safety has been renamed in security as Intrusion Modes and Effects Analysis (IMEA) (Babeshko et al. 2008).

2.2.2 Systems-Theoretic Process Analysis

Disciplinary experts are encouraged to consider security along with safety as part of Systems-Theoretic Accident Model and Processes (STAMP) (Pereira et al. 2019); the cyber security considerations in STAMP are expanded into the STPA development method for safety critical systems (Pereira et al. 2017). STPA is a safety hazard analysis process model for identifying control actions for possible hazards and accidents in causal scenarios. The hazards may be based on human and system interactions, especially human errors or mistakes (Mindermann et al. 2017).

A consistent design approach for safety and security can be based on identifying safety hazards using STPA, which may eventually lead to security concerns like vulnerabilities, threats and associated risks as is visible in STPA-Sec (Young and Leveson 2014). But the thing to be considered is that STPA involvement as safety assessment is usually at early stages of development (Pereira et al. 2017), whereas security is considered at design phase. Both are clearly not mutual, but one may lead to another.

2.2.3 Systematic Review - A STAMP Model

Usually, component failure was considered as the most common assumption behind accidents, as predicted by FMEA, fault tree and HAZOP. But with changes in technology and evolution in human roles different tools were required for safety and security analysis. With new requirements came incomplete and flawed design assumptions for system engineering. Therefore, accidents were caused due to complexities during component interactions (Lahoz 2015).

A STAMP process model revolves around examining such 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 engineer's 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 the recognition of controllers. These control actions and controllers (processes) are subsequently mapped. During design phase, 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 (Leveson 2011).

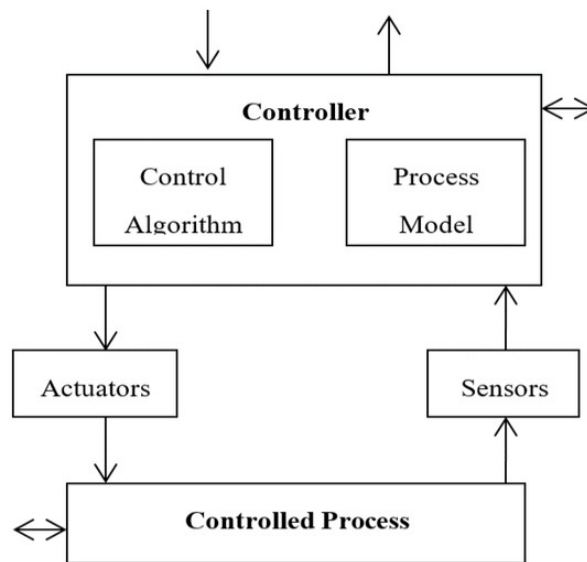


Figure 2.4: STPA Model from STAMP (Karatzas and Chassiakos 2020)

Here, the goal is to understand the behaviour of system during interaction between components. The STAMP is implemented as hazard analysis tool using STPA model as shown in Fig. 2.4 (Karatzas and Chassiakos 2020). Using STPA, system and component level requirements are dissected to identify safety constraints. These safety constraints help to understand hazard scenarios leading to violations. These violations are weaknesses or vulnerabilities in system that allow the loss (accident) to happen (Slominski 2020).

2.2.4 Applying STPA in Case Studies

The design deficiencies within critical infrastructures are timely recognised using STPA process model. The example case studies where STPA is applied are as following:

Cyber Security Case Studies by NCSC

The National Cyber Security Center (NCSC) in United Kingdom (UK) has introduced the application of STAMP/ STPA for improving risk framework for cyber security problems. The cyber security risk toolbox has been modified to include STPA approach for enterprise Information Technology (IT) infrastructure including automated/ connected products, industrial control systems and critical national infrastructure. The methodological findings from these case studies are used to inform about safety and security requirements during design. The next steps include the consideration of human factors by identifying human error sources as an impacting factor behind cyber security (Anna 2019).

Software Intensive Systems in Automotive Domain

The use of traditional safety analysis approaches for resolving safety issues for complex systems is challenging. Therefore, the use of STPA is suggested as a more detailed and comprehensive engineering approach for achieving safety goals. One such application is in automotive domain. The case study is conducted for Active Cruise Control System within BMW group of real industrial system. As a result, of this study the safety engineers were able to collaborate with software and security experts for analysis of software safety hazards and risks (Abdulkhaleq et al. 2016).

Technical Requirements for Air Force Acquisition

Past aviation mishaps were studied by applying STPA approach with a human centered analysis. This approach was used to apply Air Force acquisition process which led to safety control structures behind accidents. These accidents also provided necessary information required for identification of hazards (Summers 2018).

Another example is application of STPA for Flight Operation Quality Assurance (FOQA) for major airlines. The unsafe system behaviours were highlighted for understanding of complex component connectivities. Here, again STPA helped to recognise human behaviours in addition to design requirement flaws and issues. This study particularly pointed out the inclusion of human element for hazard scenarios (Scarinci 2017).

The Edwards Air Force Base presents an important study where flight operations were passed through safety review process using STPA method (Folse 2017).

Workplace Safety - Automate Manufacturing Process

A hazard analysis methodology using STPA is applied in automation industry to understand the safety risks from a socio-technical perspective. This perspective is evolved keeping in mind the technological advancements in factories with new methods and procedures. This motivates safety and security experts to understand the human factors

issues for determining effectiveness and efficiency for semi-automated manufacturing processes (Peper 2017).

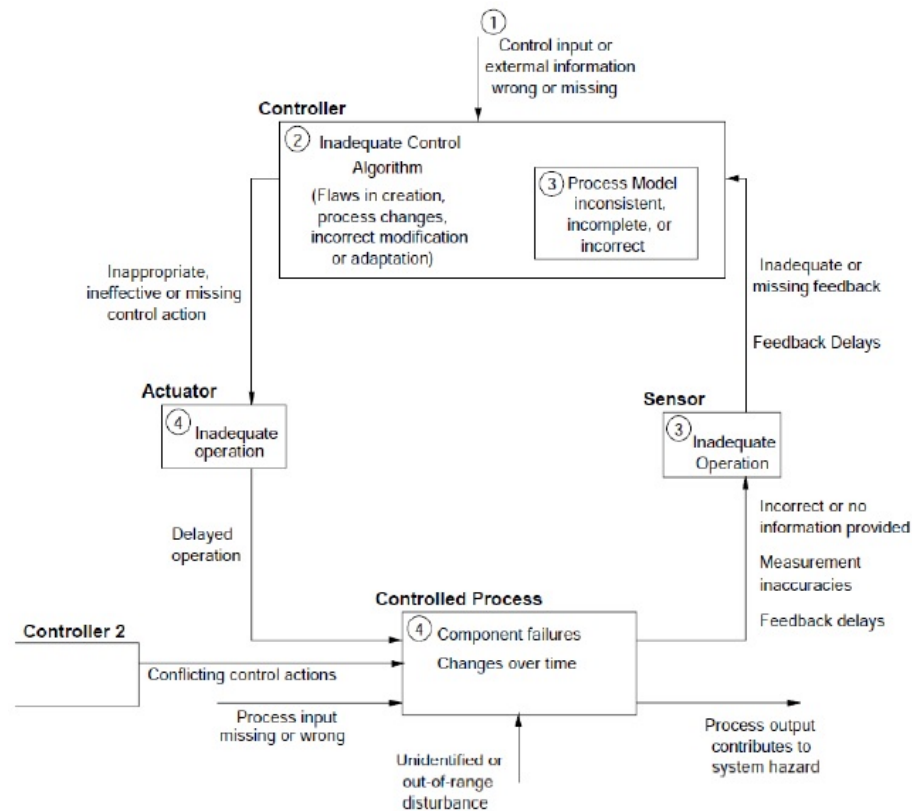


Figure 2.5: STPA Method Derivation from Control Flow Process (Karanikas 2016)

2.2.5 STPA by Safety and Human Factors Experts

Research evidence exists where the debate revolves around the use of STAMP by safety experts for safety and assurance. Usually, this safety and assurance is fulfilled by safety cases. But the case studies have proven STPA model as a far better solution (Grivicic 2019). Therefore, the safety experts are motivated to use STPA as a hazard based approach. Also, the research has shown the involvement of cognitive element as an impacting factor (Grivicic 2019).

This initiates an idea of inclusion of STPA for identifying human factors issues as a result of interaction with system. This is used to understand human error sources from human behaviour for labelling design flaws along with system hazard analysis. The goal is to integrate with human factors experts for investigating the human role. The unsafe behaviours behind system automation are used for connecting causal scenarios for hazard analysis. These causal scenarios help 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 (France 2017).

| STAMP/STPA components (Leveson N. , 2011) | Task Decomposition (Kirwan & Ainsworth, 1992) | Tabular Task Analysis (Kirwan, 1994) |
|---|--|--------------------------------------|
| Control algorithm | Description of the task, type of activity, task verb, task performance, function/purpose, sequence of activity, critical values, decisions, success criteria | Task description |
| Control action(s) | Actions, responses, speed, accuracy | Action, possible errors |
| Actuator(s) | Controls | Controls |
| Controlled process | (task) | (task) |
| Sensor(s) | Displays | Displays |
| Feedback | Feedback | Feedback |
| Process/mental model | Initiating cue/event, information | |
| Conditions and System States | Adverse conditions/states | - |
| Variables and Causal factors | Skills/training, manning, hardware location, complexity, difficulty, criticality, attention, errors | Errors |
| Other controllers | Coordination, communication | - |
| Process input(s) | Job aids | - |
| Process output(s) | Output, error consequences | Error consequences |
| External disturbances | - | - |

Figure 2.6: Task Analysis Approach Association with STPA Method (Karanikas 2016)

Upon literature survey, the control flow process for hazard analysis is used to derive STAMP/ STPA method as shown in Fig. 2.5. The significant thing about STPA is unsafe control actions where the control flow process contributes toward human error recognition. This human error source also helps to establish a link with regular task scenarios and their breakdown into description, type of activity, function, sequence of task/ action, human performance, and output. Also, TA is all about task description, action, control flow of steps, feedback which leads to identification of human error sources. Therefore, the process model of STPA can be associated with TA approach as shown by data compiled in Fig. 2.6 (Karanikas 2016). This gives security and safety experts a chance to work in conjunction with human factors experts by using classic human reliability and performance approaches for deriving better STAMP process model.

2.3 Security-by-Design & Implementation

The International Organization for Standardisation (ISO) / IEC 27002 defines Information Security as:

"Preservation of confidentiality (authorised access to information), integrity (accuracy

and completion of information and processing methods) and availability (access assurance) of information (ISO/IEC 2007)."

In addition, there are privacy properties to be ensured as well, such as authenticity, accountability, non-repudiation and reliability (Mellado et al. 2006). Here, reliability is common to safety as well. The security of a system is defined in terms of protection of its assets. *Assets* are those components of the system which need to be protected against its environment from threats and risks (Gollmann 2007). The compromise of asset happens due to an unknown vulnerability hidden beneath a system. The *vulnerability* is an exploitable weakness of the system which usually leads to threat. A design vulnerability is a logical flaw within system for example, a bug within a function or sub-system leading to threat possibilities whereas vulnerability within information system procedure or control is known as an implementation vulnerability. The *threat* is an event responsible for causing harm. The probability of occurrence of a *risk* depends on this threat and its associated vulnerability of asset in protection (ISO/IEC 2004).

In the coming sub-sections, the basics of threat modelling elements (assets, vulnerabilities, threats and risks) and their relationships are explained. This is followed by security dependency on human interaction with system in the form of potential human error using Generic Error Modelling System (GEMS) model for performance evaluation. Along with, personas and misuse cases are mentioned as design techniques for achieving security-by-design.

2.3.1 Basics of Threat Modelling

In security engineering, the identification and categorisation of threats is necessary part of risk analysis; to highlight the causes for possible incidents within an organisation including safety hazards (ISO/IEC 2007). This can be achieved by conceptualising the various threat actors or agents involved within an environment and understanding their interactions with the system. These agents or threats actors are known as *attackers*, who have the intent to conduct an *attack*. The *threat modelling* comprises of the system exploitation opportunities known as vulnerabilities as utilised by attackers and leading to threats. These threats are associated with risks causing security breaches which might have hidden safety implications and human factors issues as well.

Attack trees and misuse cases are the most appreciated techniques used to model an attacker for 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 for it to work-out accurately in threat environments (Sindre and Opdahl 2005).

In this context, the research where identification of several factors that influence the malicious behaviour of threat agents and how the system vulnerabilities are exposed

have already been conducted (Jones and Ashenden 2005). Several categories of threat actors like IT experts, students, employees, hactivist etc., have been concluded by Open Web Application Security Project (OWASP) (OWASP 2001). Similarly, the Common Attack Pattern Enumeration and Classification (CAPEC) reflect attack patterns that explain system exploitation done by the attackers (CAPEC 2007).

2.3.2 MITRE ATT&CK Framework

The MITRE Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK) framework states a bunch of threat modelling and development techniques (Corporation 2015). The framework focuses on ideology of an attacker where the execution is conducted by gaining an initial access (entry vector) by defence evasion (data manipulation) or credential access (login accounts). This leads to command and control by ex-filtration (data access). In addition, there are several matrices for understanding the target identification and attack planning. (Peters 2019). The framework allows cyber-security specialists to equip themselves with skills for analysing attack possibilities and potential attackers.

2.3.3 Potential Human Error

The security of a system is directly or indirectly dependent on human interaction (Schneier 2000). Thus, it defines security as a socio-technical work system in progress, where humans are threat to the system. The common security issues faced by the users are: procedures to complete a task, authentication required in case of multiples systems, and the theft of physical system (laptops, hard-drives etc). The survivability is an important terminology which comes into practice after the security breach have been made. The frequency and severity of an occurrence comes next (Reason 1990).

2.3.4 Generic Error Modelling System

The GEMS is used as a reference model, keeping in mind the socio-technical nature of work. The model explains the slips (failure to complete action), lapses (forgetting something) and mistakes (unintentional violation of rules) as *active failures* which are caused by humans. Even the violations made by humans is categorised as active failures. The *latent failures* are explained as the resident pathogens. They are the insiders who made the breaches. The system defects inherited due to poor design, faulty maintenance and bad management decisions impose a great security and safety threat to system. Though, the major contribution is made by the humans either intentionally or un-intentionally (Brostoff and Sasse 2001).

2.3.5 Human Performance Evaluation

The cognitive attributes and models are used to identify the human factors concerns and issues, as this is one of the determining factors for human performance and reliability (Felice and Petrillo 2011). For instance, the study in (Hammerl and Vanderhaegen 2009) offers a practicable tool for determining the cognitive attributes responsible for human performance for critical infrastructure of rail.

Previous research defines *vigilance* as the ability to remain alert for a defined period of time. Memory, attention, visual information processing abilities, auditory, and visual display are identified as vigilance increment factors, as compared to multi-tasking and reading texts which are vigilance decrement factors (Al-Shargie et al. 2019). A decision-making process that allows a user to choose best option during a given scenario is termed as *situation awareness* (Erbacher et al. 2012). During task operation, the critical thinking abilities combined with workload are necessary for better situation assessment (Golightly et al. 2009). Usually, the models for human performance tend to focus on cognitive aspects of *workload* rather than physical (Cao and Liu 2015), where this cognitive attribute is dependent on skills, Human-machine Interface (HMI) design, rules and guidelines (Hammerl and Vanderhaegen 2009). On the other hand, lack of control and fear of task failure are considered *stress* inducing factors (Conway et al. 2013). In addition, the *risk awareness* is also considered as one of the cognitive attributes and this culture is promoted by expertise, technical abilities, better communication skills and knowledge (Jen 2012).

2.3.6 Design Techniques

Security-by-design need for secure foundation is dependent on the design techniques and processes chosen during its life-cycle. As security is an ongoing process and not an end product. The personas and misuse cases are good design techniques for assuming potential security breaches.

Personas and Attacker Personas

Personas explain the archetypical behaviour of users. This is based on ground information collected from similar environments, where the user is expected to act (Cooper 1999). According to (Norman 2004), 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 design concepts have been told by identifying four categories of personas (Soegaard and Dam 2013):

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.

In addition to roles, personas can also be supported by stories and scenarios. A better and refined system view can be obtained by generating personas within relevant narrative scenarios and real-life situations (Nielsen 2013). The story-based personas have better chances of explaining the user behaviors (Pruitt and Grudin 2003). This way the personas can be utilised to explain different contexts of system and environments in which they are operating. A persona built from a user-centered design approach has better chances of being used for various analysis purposes (Faily and Fléchais 2010) for example, 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 are used to visualise the problem space in which there is risk of compromise of security (Atzeni et al. 2011). From an attacker's view, the possible threats faced by the system based on system vulnerabilities which are otherwise not visible.

Misuse and Misusability Case

The traditional use-case approach is used to write narratives for misuse cases, for identifying security requirements Sindre and Opdahl (2005). Just as a use-case comprises of use-case and actors, a misuse case comprises of misuse case and misusers. Misusers are possible actors of a system with mischievous intent to interact within system. Due to which a detailed security threat analysis of system can be conducted by security experts (Sindre and Opdahl 2005). However, the lack of appropriate principles and guidelines for writing a use-case, makes it an approach with open-end results and solutions.

The misuse cases evolves into misusability cases, which are the design solutions for security along with usability of a system (Faily and Fléchais 2016). The risk scenarios during threat modelling are explained using misusability cases. These help to determine the usability concerns that affect the security decisions during design of a system.

2.4 Security leading to Human Factors Engineering

The threat to a system in an environment is usually caused by an attacker which is the human element responsible for compromising the security (Schneier 2000). This identifies humans as the biggest source for human error (Reason 1990). Similarly, the security engineers now give importance to human dimension of system during design phase by considering the usability attributes during asset identification, threat scenario, misuse case, task duration, responsibility modelling etc (Faily and Fléchais 2010). Therefore, the concept of effective information security revolves around the idea of Human Computer Interaction - Security (HCI-security) of the system. The HCI-security experts use design principles and user-centered approaches for designing usable security (Shostack 2014).

In the coming sub-sections, the secure and usable modelling techniques along with available tool-support options are explained:

2.4.1 Toulmin's Model of Argumentation

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 validity, *warrants* that act as inference rules connecting the grounds to the characteristic, and *rebuttals* that act as counter-arguments for the characteristic. A model qualifier is also used to describe the confidence in the validity of the characteristic.

These argumentation models are used to act as the source of confirmation, for data sources used as document references for designing security approaches like roles and personas definition. These document references are known as *factoids* which are facts in the form of statements acting as reliable information. Each factoid is gathered after carefully scanning document references. The document references in the form of factoids (arguments) are elicited by carefully reading the data sources, which are used to do the affinity diagramming. For this purpose, *Trello* board is used to organise the factoids into different groups. The assumption data is organised into clusters of similar characteristics in several sessions and discussions with relevant stakeholders.

2.4.2 Use-Case Scenario and Template

The inclusion of goal and responsibility in single structural format is represented as a use-case (Cockburn and Bank 1997). Usually, use-case is written in the form of scenario where an actor is associated with goal leading to fulfilment of responsibility. A general template comprises of use-case name, scope, level, pre and post conditions, actions, and other characteristics enabling to consider functional requirements and scope of project (Cockburn 26-October- 1998).

2.4.3 KAOS - Goal Modelling Language

Goal and task models can help security engineers to better understand the system threat model. The KAOS is a method for analysing, specifying, and structuring goals required for a system (Dardenne et al. 1993). The goals and tasks modelled using Unified Modelling Language (UML)-class diagrams, may allude to the security requirements that need to be fulfilled, along with possible obstacles that model obstructions to system goals.

2.4.4 Integrating Requirements and Information Security

The IRIS process framework (Faily 2018) was devised to understand how design concepts associated with security, usability, and software engineering could be aligned. It is complemented by the 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 (Faily and Flechais 2011).

The IRIS framework is based on a meta-model with six views of a system being designed or examined:

Environment: The context in which the system is supposed to function, and where all the objects and subjects are defined.

Asset: The object present in a particular context that needs to be protected or safeguarded.

Task: The assumptions made about the behavioural specifications of users involved and how they are supposed to interact with the system.

Goal: The objective to be achieved by the users and the system's ability to satisfy the users.

Risk: The probability of occurrence of system risks based on threats and vulnerabilities.

Responsibility: The subjects interacting with system based on their defined roles.

The core IRIS concepts are shown in Fig. 4.2. 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 designed with safety in mind, IRIS provides a foundation for integrating safety, security and human factors engineering.

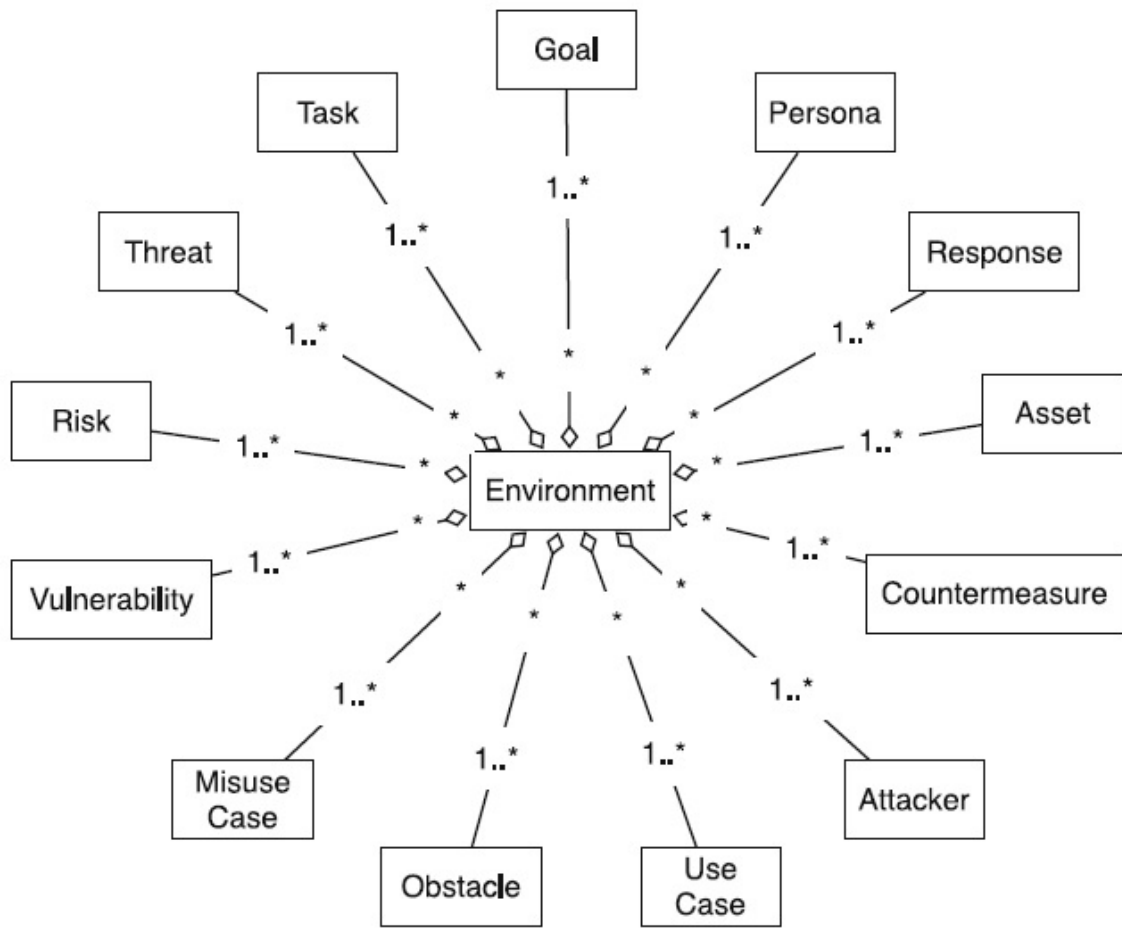


Figure 2.7: IRIS Framework Meta-Model Views (Faily 2018)

2.5 Human Factors Design Solutions

The reduction of negative aspects and enhancement of the positive aspects within a system in terms of user experience is defined as interaction design (Sharp et al. 2007). Interactive products require knowledge of users who are going to use systems. For this purpose, tried and tested user techniques should be utilised. The way people communicate and interact define the goals for interaction design. Interaction design is a wide open umbrella under which, system design, requirements engineering, computer science, human factors, and Human Computer Interaction (HCI) all falls.

The ISO/IEC 9241 as Ergonomics of human-system interaction standard defines usability as:

"The effectiveness (accuracy and completeness), efficiency (goals achieved with respect to resources) and satisfaction (comfort and acceptability) with which specified users achieve specified goals in particular environments (ISO/IEC 2010)."

The human factors and usability designers need to understand the users, technolo-

gies, interactions, emotions and human experiences (Sharp et al. 2007). Every team member has a different perspective and they need to include design principles for implementing user-centered design approach for TA (Preece et al. 2002), as explained in up-coming sub-sections.

2.5.1 User-Centered Design Approach

The establishment of requirements for providing design solutions from the perspective of desired users is defined as a user-centered design approach. The approach is implemented as an iterative process and the desired users needs and requirements are termed as usability goals. Even the evaluation of system design is based on these usability goals (Preece et al. 2002). The storyboards, affinity diagramming, mental models, personas, scenarios, use-cases, workflow diagrams, prototypes, stakeholder interviews and surveys are few of the many techniques used for adopting user-centered design approach (Mastery 2012).

2.5.2 Fundamental Design Principles

The following fundamental design principles need to be considered while implementing a user-centered design approach (Sharp et al. 2007):

Visibility: The controls to be performed as part of user task should be clearly visible to user, with appropriate lighting.

Feedback: The users should be provided with timely outputs and results to ensure the completion of jobs and tasks.

Natural Mapping: The controls for tasks and their outputs should have a clear relationship, helping the user to develop mapping of concepts.

Affordance: The interface for tasks should be easy to understand and use.

Constraint: The user task design should limit the possibilities for making a mistake or an error.

Convention: Easy to learn conventions should be adopted for users.

Environment: The usability attributes of an environment should be considered, in which the user task will be performed.

Work flow: The tasks to be performed by users to achieve system and user level goals.

Work load: The mental (cognitive) or physical load placed by tasks on users during performance should be considered.

These design principles need to be made part of specifications during design decisions. By design those specifications of the system are meant which are responsible for determining the process required for product development. Similarly, design specifications for any system can be improved to minimise the chances of any human errors or mistakes, by conducting a detailed TA exercise. The human factors experts have suggested TA as one of the many approaches for making design improvements (Affairs 2013).

2.5.3 Task Analysis Processes and Tools

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 (Diaper and Stanton 2004). TA is used to determine 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 (Affairs 2013).

There are two main types of TA: hierarchical and cognitive task analysis (Diaper and Stanton 2004). The HTA is conducted to determine the hierarchy of tasks by decomposing high-level into low-level tasks (Crandall et al. 2006). The CTA focuses on the cognitive load put by tasks on users depending on their cognitive abilities (Hammerl and Vanderhaegen 2009). The most notable techniques used for eliciting data for TA are: interviews, focus group discussions, surveys, workshops, and questionnaires.

Table 2.1: Methods and Tools for Task Analysis with Applications

| <i>Task Analysis Method</i> | <i>Tool-Support</i> | <i>Application</i> |
|---|--|--|
| Hierarchical Task Analysis (HTA) (Embrey and Zaed 2021) | Human Factors Risk Manager (HFRM), Human Factors Workbench (HFW) | Risk Scoring, Failure Mode, Error Description |
| Cognitive Task Analysis (CTA) (Militello and Hutton 1998) | Applied Cognitive Task Analysis (ACTA) | Cognitive Demand & Skill, Training Recommendation, Interface Improvement |
| Ecological Task Analysis (ETA) (Davis and Burton 1991) | - | Control Theory, Cognitive Psychology |
| Operator Action Event Tree (OAET) (Embrey 2000) | Event Tree (Success & Failure) | Human Reliability Assessment |
| Flow Diagram (Embrey 2000) | Flow Chart | Binary Decision Logic |
| Influence Modelling and Assessment System (IMAS) (Embrey 2000) | Cause-Consequence Model | Skills Diagnostic, Mental Model |
| Critical Action and Decision Evaluation Technique (CADET) (Embrey 2000) | Critical Action or Decision (CAD) | Potential Cognitive Error, Failure Scenario |

The decisions about design, training needs, human error analysis, stress and workload management are dependent on TA (Embrey 2000). The human factors experts aim to identify human error sources for resolving human factors issues. As these human error sources are considered determining factors for risk and safety analysis during accident investigations (Embrey and Zaed 2021). Also, the Training Needs Analysis (TNA) and

mental workload behind tasks is analysed to identify the training gaps in order to train operators interacting with a system.

The TA approaches are used by human factors experts to identify the system design and engineering requirements. An application using software tools for error identification, training requirements and task load using TA is presented by Human Factors Workbench (HFW) where Predictive Human Error Analysis (PHEA) and Performance Influencing Factors (PIFs) analysis are among notable tools (Embrey and Zaed 2021). Also, for HTA the automated tool-support is provided by Human Factors Risk Manager (HFRM), where risk scoring, failure modes, and error descriptions are applied for TA (Embrey and Zaed 2021).

Usually, Computer Aided Software Engineering (CASE) tools and components are used for the representation purposes, including UML, scenario-based design and Concur Task Trees (CTT) (Diaper and Stanton 2004). The UML pre-defined specification formats in the form of use-cases are used to include description of actor/s, specific conditions, steps and exceptions for TA, but is limited to data representation. CTT enables to understand the hierarchical task breakdown, representation of activities using graphical syntax, and task allocation including attributes, but it lacks an understanding of cognitive attributes (i.e. mental workload) needed for accomplishment of tasks.

A brief summary of TA approaches and methodologies as supported by available software tools along with their applications is shown in Table. 2.1. Although, the point of consideration lies beneath the choice of appropriate method depending on desired application. For example, CTA is applied for determining cognitive demand and skill, whereas HTA is more suitable for risk scoring and error description.

2.6 Overlap between Human Factors & Safety Engineering

The tendency of humans to make errors during their interaction with systems, led to recognition of human factors engineering. The safety of people in critical infrastructures like rail is often compromised due to this occurrence of human error (Baysari et al. 2008), as also acknowledged by (O'Hare 2001). The safety hazards identified by utilising secure and usable IRIS framework (Faily 2018), has provided a strong linkage with human errors as well.

The identification of human error during design of safety critical systems should be the top 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) (CEN 2017). The validation of this aspect was made by (Kirwan 1997), where the risk assessment for design of safety of systems like transportation industry was done by considering the Human Reliability Analysis approach. The latent failures originates from active failures and usually have same catastrophic effects on human life (Reason 1990). Due to the complexity of consequences of incidents, there is no

well-defined methodology for determining the sources of these active and latent failures (Shorrock 2007).

2.6.1 Swiss Cheese Model of Accident Causation

According to the Swiss Cheese Model of accident causation (Reason 1990), there should be multiple layers of defence within a system or an organisation against the emergent errors or mistakes, which may eventually lead to hazardous accidents. As the name indicates, the model takes the inspiration from a slice of cheese as shown in Fig. 2.8, where the holes represent the human weaknesses and different slices act as the barriers. Some of the holes are active failures whereas some are latent failures, and all the holes have to be aligned to each other at the same time for the accident to occur.

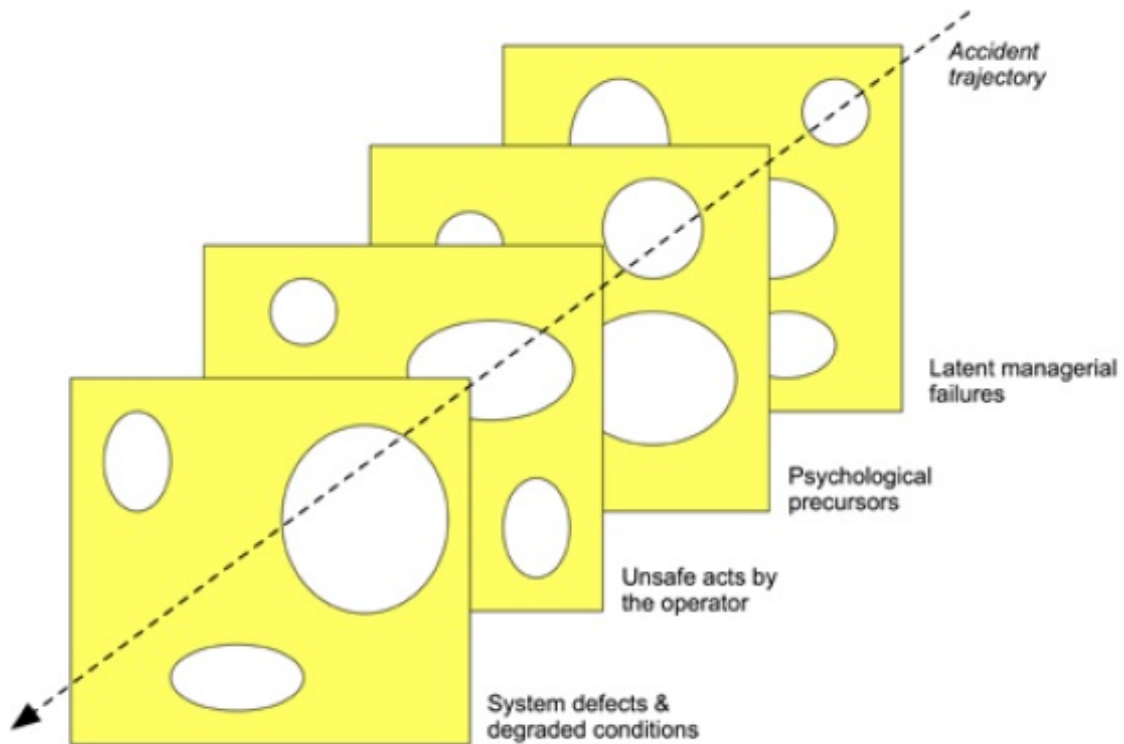


Figure 2.8: Swiss Cheese Model of Accident Causation (Reason 1990)

The *human* is the most important aspect of this model, whose intent and capabilities are usually ambiguous. Therefore, not all possible holes can be generalised before time. Based on Reason's error taxonomy (Reason 1990) of cognitive, behavioural, personal and organisational factors, the HFACS framework represents four levels of failures and error sources (Wiegmann and Shappell 2003).

2.6.2 Human Factors Analysis and Classification System

The HFACS framework as shown in Fig. 2.9, represents four levels of failures (error-sources) by providing a multi-level categorisation as follows (Zhou and Lei 2018):

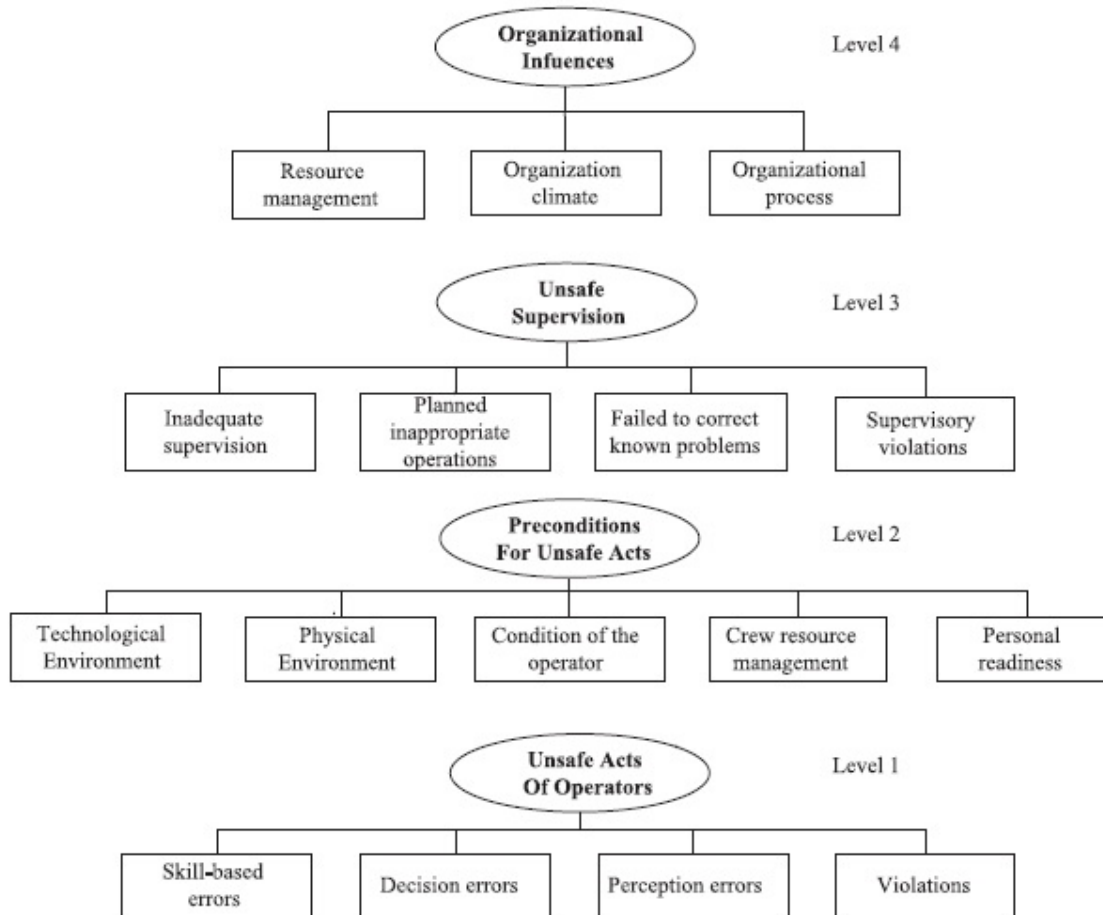


Figure 2.9: HFACS Framework (Zhou and Lei 2018)

Unsafe Acts of Operations: Used to identify the skill-based errors, decision errors, perception errors and violations made during operations.

Preconditions for Unsafe Acts: Several circumstances may contribute like poor design of equipment and controls in technological environment, uncontrollable and unpredicted physical environment (weather), mental state of operator, poor communication during crew resource management, poor safety awareness and insufficient staff trainings.

Unsafe Supervision: The errors and mistakes as a result of: inadequate supervision, inappropriate operations, failure to correct known problems and supervisory violations.

Organisational Influences: Resource management, organisational climate and oper-

ational process (regulations) are identified as the biggest sources behind active failures.

The HFACS have been used by rail stakeholders to determine the human error sources behind accidents and incidents (Zhou and Lei 2018). Human factors experts use this framework to investigate the accidents by identifying and classifying the human causes in the form of errors, mistakes or violations. Eventually, it is the job of the system design to ensure safe acts, by making certain that there is no room for any human mistakes or errors. However, to date, there has been no work on how it can be used to consider safety or security attributes of rail system.

2.7 Rail Infrastructure Design and Evaluation

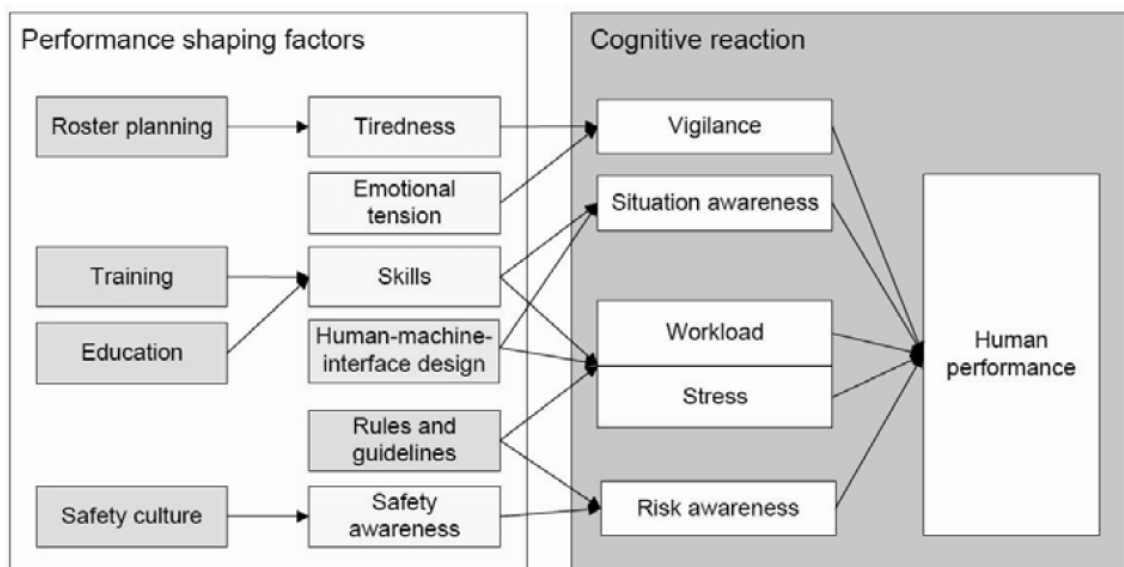


Figure 2.10: Cognitive Reactions Responsible for Human Performance (Hammerl and Vanderhaegen 2009)

The rail infrastructure has long been managed in accordance with health and safety standards, working within legislative requirements such as in UK the Railway Act 2005, under guidance and supervision from bodies such as the Railway Safety and Standards Board (RSSB) and Office of Rail Regulation (ORR). More recently, the shift to digitalisation stipulated by the ERTMS imposed by European Union (EU) 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 (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 et al. 2016). 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 (Bloomfield et al. 2018).

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 (Alper and Karsh 2009) highlights the value of combining safety with human factors.

Rail infrastructure should be resilient enough to block opportunities of human error, but not at the cost of security and safety. For example, work by (Cacciabue 2005) describes how the HERMES risk management approach assesses the chance of human error, but while it identifies safety and reliability components required in rail, it does not consider cyber security.

Similarly, the Generic Risk Assessment Log presented by Randstad Rail mentions all the possible events that can lead to safety hazards, but does not mention the associated security concerns. The 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 (Zhou and Lei 2018). The framework does not mention any dependency towards emerging security and safety concerns.

2.7.1 Human Performance and Reliability

According to the RSSB in Great Britain, the Human Factors Working Group has identified system design, user training, organisational culture, and cognitive reactions as the evaluators for human performance. Nowadays, due to technological advancements in rail infrastructure the tasks are becoming more centered around mental (cognitive) abilities as compared to physical.

Also, with the implementation of ERTMS, the working relationship is more dependent on team coordination capabilities for example, the train driver and signaller work in conjunction with each other to ensure safe and efficient operations. Thus, cognitive attributes and models are used to identify the human factors concerns and issues, as this is one of the determining factors for human performance and reliability (Felice and Petrillo 2011). For instance, there is a study which offers a practicable tool for determining the cognitive attributes responsible for human performance in rail for the roles of train driver and signaller as shown in Fig. 2.10 (Hammerl and Vanderhaegen 2009).

2.7.2 Tools and Resources

The following are available tools and resources in use by rail:

Rail Risk Toolkit

The RSSB enlists a range of tools to be used for achieving safety and human factors goals. The most common among which is 'Human Factors Toolkit'. The toolkit provides documentation on human performance dependent on factors like design, training, staffing, culture, and conditions. Several approaches such as cognitive task analysis, critical decision method, design scenario analysis, fault trees, hierarchical task analysis, interviews, questionnaire, situation awareness, system usability scale, team cognitive task analysis among others, are available for performing analysis for each of these factors. The toolkit aims at covering all the latest trends in safety and human factors background (RSSB 2019).

The RSSB toolkit is not software based and it enlists approaches to be specifically used by human factors experts whereas the ERTMS digitalisation requires several safety and security concerns to be resolved as well. Especially, when human factors approaches like human error assessment and reduction, HAZOP, fault trees, and human reliability analysis are mentioned but their utilisation by human factors experts or subject matter experts is not clearly defined.

Adelard Tool-Support

Adelard has provided a list of approaches to be implemented as tool-support for management of safety. The Assurance and Safety Case Environment (ACSE) provides an easy assurance of safety for reducing the possible risk factors. The approach is built on Claims, Arguments and Evidence (CAE) model of presentation and aims to provide effective safety arguments. Along with, the Goal Structuring Notation (GSN) is used as an argumentation notation. The approach is responsible for providing effective safety cases for security-informed safety (Adelard 2019).

Though, the Adelard tool-support is built for safety experts only, but it helps by giving evidence for two arguments. First, the CAE model of presentation and GSN notation augments the secure and usable modelling relationship, which is developed in Section 2.4 as well. This devise an argument that task and goal modelling techniques can be used to determine safety goals as well. Second, the security-informed safety is extracted from safety cases. Thus, it confirms that safety and security are dependent on each other and similar approaches are being adopted by rail stakeholders.

International Union of Railways - Safety Control

The International Union of Railways (UIC) aims to make its system more resilient against emerging safety and human factors concerns. Their Human Factors Working Group members are integrating human factors with system safety with appropriate tools (UIC

2019). Here, again the tools are not software based and the experts are reliant on documentation based interpretation of approaches. Also, despite a clear vision and observed relationship between human factors (human error) and security, the integration of safety and human factors lacks security design.

Table 2.2: Brief Comparison of Standards and Practices in Rail

| Standard/Practice | Safety | Security | Human Factors |
|---|--|-----------------------------------|-----------------------------|
| Railway Act 2005 | Health and Safety Standard | - | - |
| Railway Safety and Standards Board (RSSB) | Rail Risk and Safety Management | - | Human Factors Toolkit |
| Office of Rail Regulation (ORR) | Health and Safety Standard | - | - |
| Common Safety Method for Risk Evaluation and Assessment (CSM-REA) | UK Specific Safety Concepts | - | - |
| As Low As Reasonably Practicable (ALARP) | Safety Risk Management | - | - |
| Codes of Practice (CoPs) | Safety Design | Security Design | - |
| Human Error Risk Management for Engineering Systems (HERMES) | - | - | Human Error Assessment |
| Adelard Tool-Support | Assurance and Safety case Environment (ACSE) | Security-Informed Safety | - |
| International Union of Railways (UIC) | Safety Control | - | Human Factors Working Group |
| ISO 27001, 27002 and 27005 | - | Information Security | - |
| NIS Directive Corporation Group Guidelines | - | Cybersecurity Measures | - |
| IEC 62443 Standard | - | Mitigate Security Vulnerabilities | - |
| Technical Specification (TS) 50701 | - | OT-Specific Security Measures | - |
| CYRail Recommendations | - | Cybersecurity Assessment | - |

2.7.3 Brief Comparison of Standards and Practices in Rail

The Table. 2.2 shows a brief comparison of safety, security and human factors design concerns among different standards and practices in rail. According to this table, the safety considerations are common to almost every practice against security and human

factors. Also, there is not even a single standard, practice or tool available which is compliant of all three design factors which are safety, security and human factors.

2.8 Summary

There have been some common grounds for safety and security engineering culture by their experts (Section 2.2). Here, several process-techniques and approaches contribute towards the identification of potential safety hazards and security risks. The foundational approach is found to be STPA (Section 2.2.2) where the safety hazards are analysed and linked up with human factors approach such as TA for better human performance evaluation (Section 2.2.5). The STPA is found to be a process model which connects with security as well. As the source of security risks is found to be potential safety hazards under analysis by STPA.

The evidence also shows the consideration of safety aspects by human factors specialists (Section 2.6). One such example is of HFACS framework which helps to identify and categorise human error sources (Section 2.6.2). These human error sources have the tendency to cause safety hazards leading to catastrophic consequences (accidents).

Along with, there have been strong dependency towards human factors by security engineers (Section 2.4). The IRIS framework with CAIRIS tool-support helps to visualise this link by highlighting the security aspects (asset, vulnerability, threat, risk, obstacles etc) and human factors aspects (role, personas, task, use-case, goal etc) as well (Section 2.4.4). Using IRIS, multiple views of system in the form of environment, asset, task, goal, and risk can be put forward for analysis. This gives both security (risk analysis) and human factors experts (usability and cognitive attributes) to contribute together.

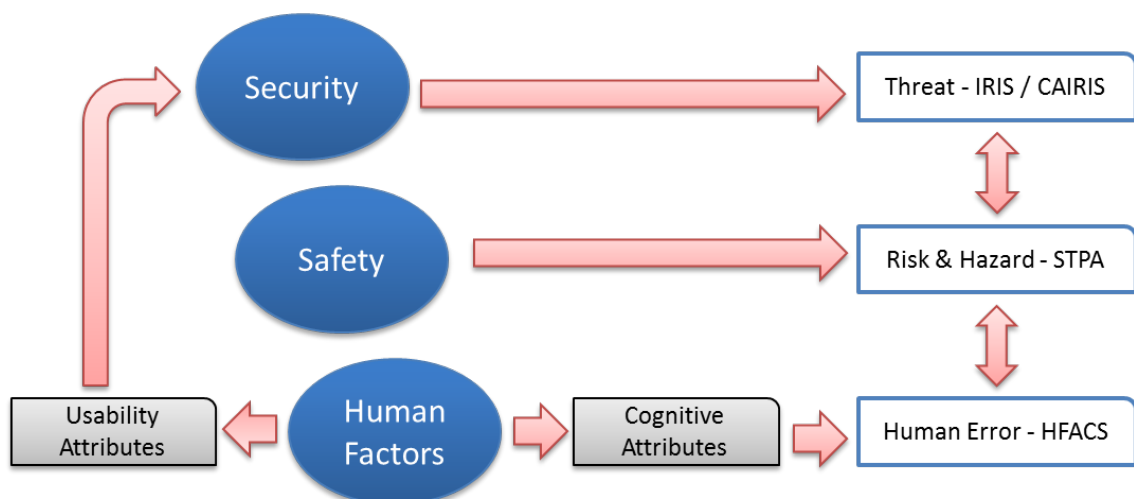


Figure 2.11: Security, Safety and Human Factors Existing Approaches

Thus, on the basis of available knowledge, a basic idea about the linkage between

existing approaches which are utilised well as best practices by respective practitioners is concluded with the help of Fig. 2.11.

The security along with usability attributes from human factors guide towards the identification of threats which consists of basic threat modelling elements namely assets, roles, personas, vulnerabilities, threats, tasks, goals, and obstacles. The IRIS framework explains the relationships and dependencies between these threat modelling elements. Moreover, the CAIRIS tool-support can be utilised to visualise these relationships in the form of models. These threats lead to the identification of potential risks and safety hazards. Here, STPA can be used to classify hazards in detail including human error sources and determine the control actions to be taken against these identified hazards. Sometimes, the human error is responsible for compromising the safety of an environment. These human error sources are determined and classified using HFACS framework.

The tendency of humans to make errors or mistakes is dependent on cognitive attributes like vigilance, situation awareness, workload, stress, and risk awareness. These cognitive attributes are found to be responsible for effecting the performance of human, which can be labelled as human factors concern. The threats, risks, hazards, and human error sources are derived from different disciplines but are co-related to each other. This co-relation is responsible for defining scope for an integration between safety, security and human factors engineering.

Chapter 3

Methodology

In this chapter, an overview of available research approaches with respect to safety, security and human factors engineering is given. This leads to perspective consideration for proposed research methods. Based on this, a detailed explanation of research model adopted for this thesis is mentioned including, literature and systematic review, theoretical cyber-model framework, interviews, and qualitative case study analysis for validation of safe, secure and usable design framework.

3.1 Research Approaches

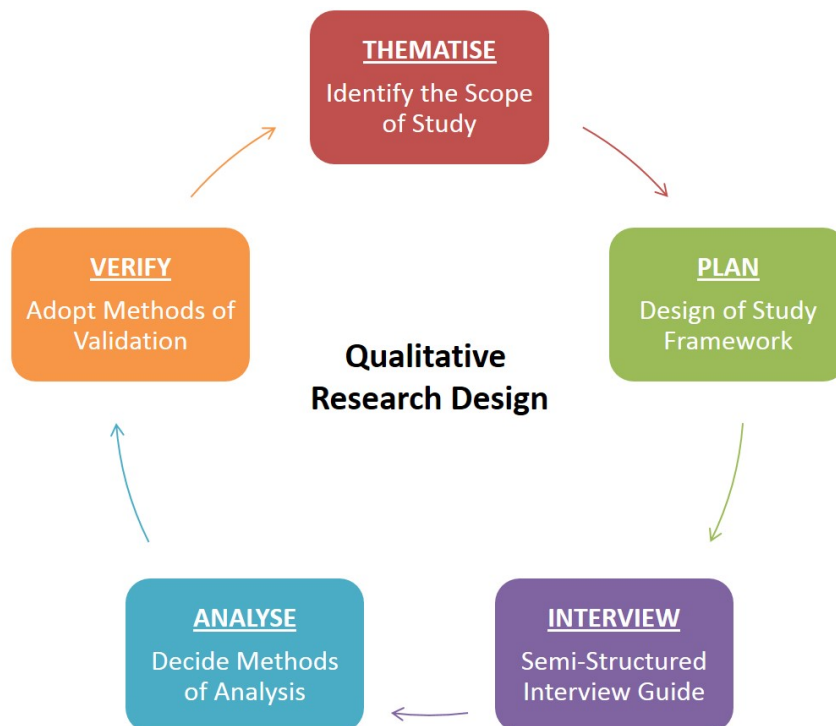


Figure 3.1: Qualitative Research Design Cycle (Steeves 2018)

Among quantitative, qualitative and mixed research approaches, the *qualitative research* approach is adopted for this PhD. The choice between qualitative and quantitative research approaches is dependent on following factors:

Deep Understanding of Disciplines: During this research, three engineering fields are understood from an over-lapping point of view. These three disciplines are safety, security and human factors engineering. These disciplines are wider in scope and finding an intersection concept between all three is the major challenge. For this purpose, the qualitative research approach is better suited for developing a unique depth of understanding which is difficult to obtain from a quantitative approach (fixed survey). For example, using qualitative research the concept of human error (human factors) is dissected to understand human potential, performance evaluation and behavior. This is used to link potential of humans to make mistakes (TA) with safety (hazards) and security (risks) aspects of research using interviews, documents and artefacts.

Opportunity to Elaborate: The quantitative research approach is based on facts and figures using statistical analysis. This leads to descriptive data with interpretation differences. In order to focus on 'why' part of research (interpretation), the qualitative research methods are better suited such as case study application, interviews etc. For example, by applying a theoretical framework to a case study gives a chance to study the results and analyse them with respect to stakeholders (participants). This gives a chance to even improve using feedback and review.

Flexibility: A standardised set of procedures and statistics are involved in quantitative research which does not leaves any room for un-foreseen situations and circumstances. However, a qualitative research allows flexibility according to present situation. For example, during pandemic (Covid-19) the interviews were scheduled with signallers from *Network Rail*. But due to staff shortage, new participants were to be arranged. This involved managing constraints such as rescheduling of interviews, time-line management for PhD, use of online platforms (virtual), and online feedback.

Generally, the qualitative research design cycles around, identifying the scope of study by *thematizing*, *planning* the design of study framework, conducting structured/semi-structured *interviews* for data collection, deciding methods of analysis for *analysing* results, and adapting appropriate methods for *validation* as shown in Fig. 3.1 (Steeves 2018).

In the following sub-sections, a brief overview of approaches involved with respect to development stages is explained.

3.1.1 Establishing Requirements

The research identifies five key issues during data gathering and establishment of requirements. It begins by setting up goals and identifying desired participants, relationship with targeted participants, triangulation (combination of qualitative and quantitative research methods) and pilot studies (Sharp et al. 2007). The safety and security experts depend on establishing requirements for setting up safety goals. Also, during verification and validation these requirement specifications are consulted by stakeholders.

Another point of focus is data recording, which can be done in the form of notes, audio and video. The unstructured and structured interviews plus focus group discussions, questionnaires (checkboxes and Likert scale), and observational studies (ethnography) are all examples of some best practices. Usually, quantitative (statistics, mean, median, mode, graphs, histograms and pie-charts) and qualitative (recurring patterns, themes, data categorisation and critical incidents) analysis is done either independently or as a combined approach for collecting the requirements during research (Sharp et al. 2007).

3.1.2 Design and Prototyping

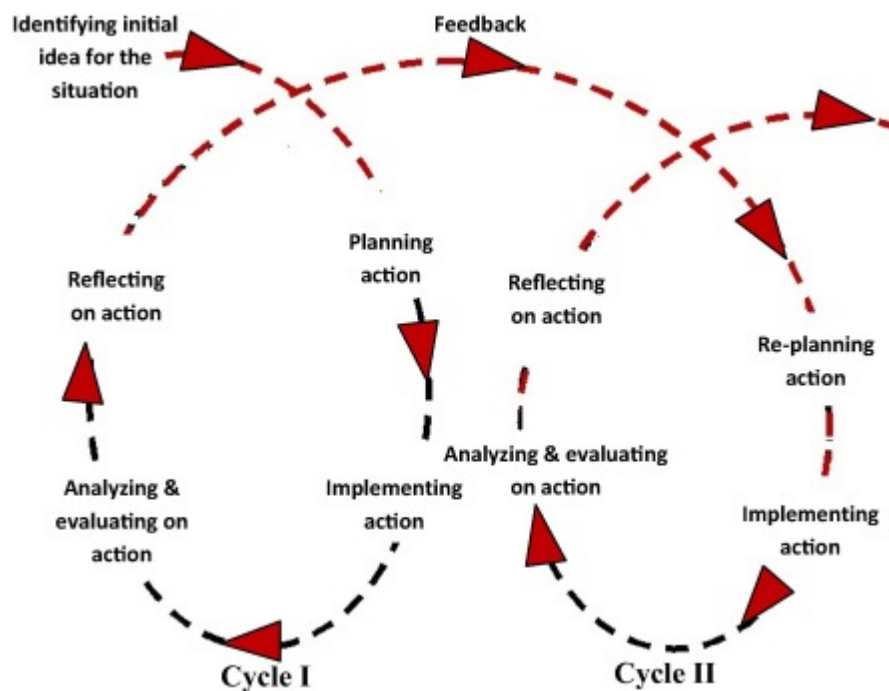


Figure 3.2: Action Research and Grounded Theory Cycle (Chen and Cheng 2015)

Grounded Theory with open coding, axial coding and selective coding along with *Action Research* is a good framework for design and analysis as shown in Fig. 3.2 (Chen and Cheng 2015). The cycle revolves around identifying ideas and planning action points, implementing those actions points, analysing, and evaluating by reflecting on action points.

The feedback from this cycle is used to inform re-planning of action points for next cycle, and the process continues. Another process is *Distributed Cognition Analysis*, where the domain is described at various levels of granularity. The *Activity Theory* is where three layers of interest are: operation and condition, action and goal, and activity and motive.

The four approaches to design process are: user-centered design, activity-centered design, system design, and genius design (Chen and Cheng 2015). The requirements are gathered, the initial design is built, the prototyping (low-fidelity and high-fidelity) is done and evaluation is made. The interaction design framework is summarised as *DECIDE*, where D is to determine the goals, E is to explore the problem, C is to choose the evaluation method, I is to identify practical issues, D is to decide how to deal with ethical issues, and E is to evaluate, analyse, interpret, and represent data (Sharp et al. 2007). The human factors experts deeply rely on these qualitative research approaches for achieving their usability goals.

3.1.3 Theoretical and Experimental Approaches

Sometimes, theoretical and experimental approaches are used by security experts for answering the scientific questions. The approach entails, determining the problem, followed by an identified solution, and then using test data scenarios to validate the outcomes by performing experiments (Maxion et al. 2010). Here, the theoretical data and findings are used to predict the expected experimental results. *Action Research* lays theoretical foundation for qualitative analysis by providing an understanding about the research plan from expected activities as input and results as output (Rearick and Feldman 1999).

3.2 Proposed Research

The proposed research methodology needs to consider following perspectives:

- The research needs to provide foundational grounds based on valid artefacts where safety, security and human factors engineering concepts can be integrated together for their respective practitioners.
- The research output should be a theoretical design framework, which will contribute knowledge. This design framework will be used to solve real world problems in rail infrastructure where the safety, security and human factors concerns are in tension.

3.3 Application of Research Methods

The following research methods are adopted for this research project as shown in Fig. 3.3:

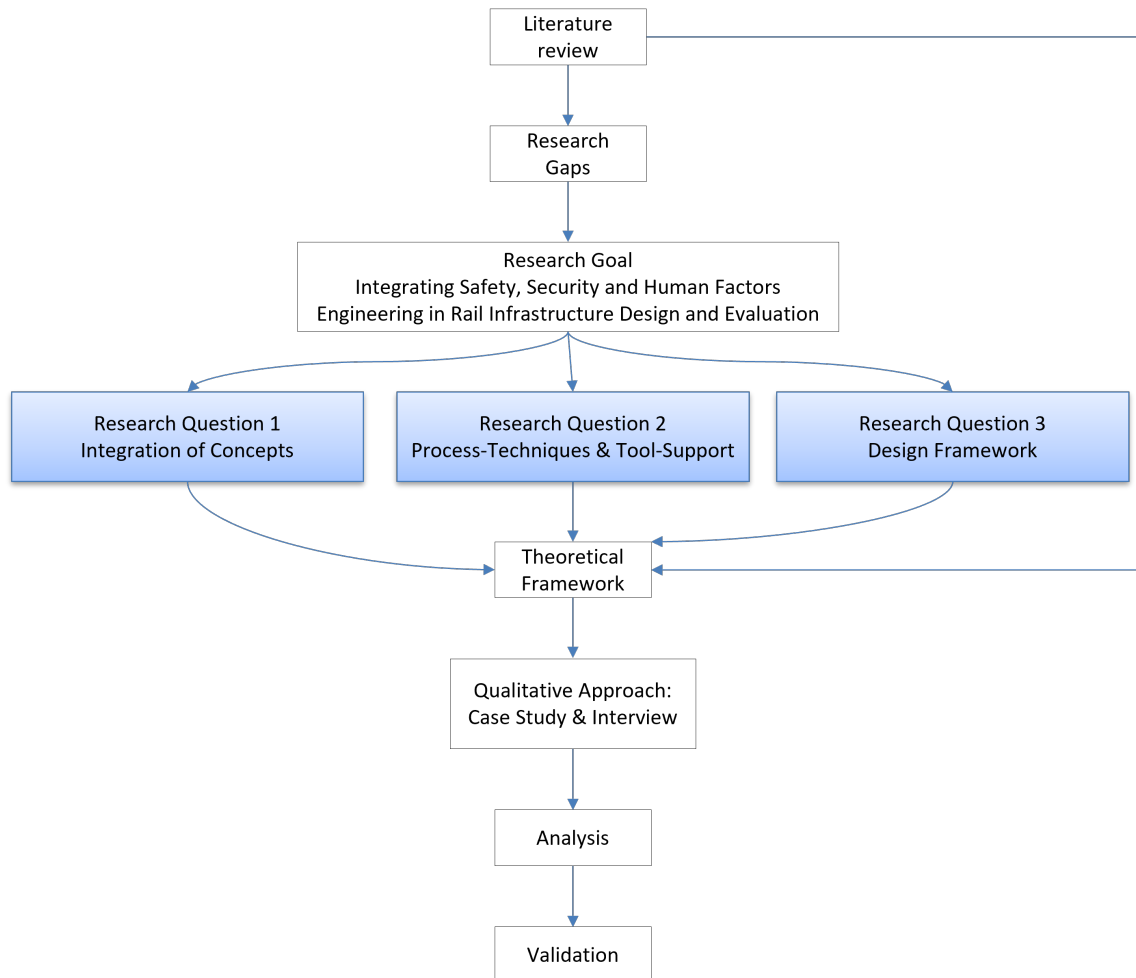


Figure 3.3: Application of Research Methods for Thesis

3.3.1 Literature and Systematic Review

The research planning begins by narrowing down terminologies for research, where hypothesis is developed. This literature search is used to identify research questions, by keeping in mind the research proposal as shown in Fig. 3.4. This PhD research is focused around three domains: safety, security and human factors engineering. For this purpose, the literature review of related work is conducted in Chapter 2, in order to build the foundational grounds for design framework. The literature about safety critical-systems, security-by-design and human factors engineering is read to develop an understanding for individual domains.

During this literature and systematic review, the google scholar, springer, and IEEE Xplore etc are used as tools and databases. The Journal Impact Factor (IF) is monitored when short-listing research articles for review. The keywords such as safety engineering, safety and security, security engineering, security and human factors, safety and human factors, and human factors in rail are used during research. In order to make research

more focused, the field-centric process-techniques such as risk analysis, potential hazards in rail, human error sources, task analysis, threat modelling, asset associations, vulnerability identification, use-case specifications, STPA analysis, and rail standards etc are studied and explored.

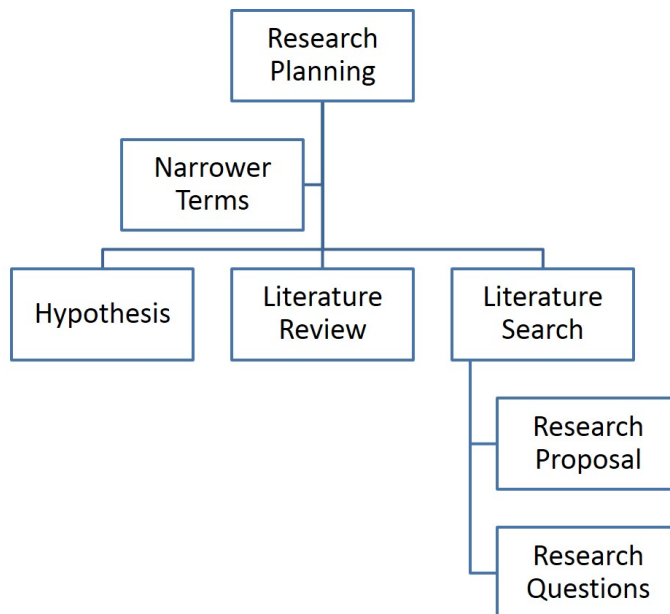


Figure 3.4: Literature and Systematic Review Process

Here, the approaches and process-techniques used by these engineering fields is well understood. The existing knowledge of tool-support is also gathered. The identified overlap between these domains is used to build a systematic review, where the assumptions about new relationships between safety and security, security and human factors, and human factors and safety are made based on available facts. These assumptions are responsible for validating artefacts, for setting up the integration stage for safety, security and human factors engineering. In addition, a detailed review about the rail infrastructure including their operational contexts, applicable standards and practices are also conducted.

3.3.2 Theoretical Cyber-Model Framework

For cyber security and safety analysis, theoretical cyber-models are generated to understand the identified problem in detail. The methodology begins by identifying the appropriate question of interest, then the gathered data by literature review or any observational study is applied to generate a visual model (W. Edgar and O. Manz 2017).

For this research, theoretical cyber-models consisted of asset, vulnerability, threat, risk, task, goal, and obstacle as basic elements. The choice of these elements is based on taxonomy of threat modelling, risk and safety analysis. The already available tool-

support is determined and open-source CAIRIS is prioritised due to its secure and usable attributes. This methodology is an on-going process and elements are filled in the form of use-case specifications during TA using CTA and HTA models. Also, STPA is conducted using KAOS and DFD for control structures and control flows. No order of modelling or processing of data is followed. Using this theoretical cyber-model, an understanding about the existing relationships and dependencies between the elements of threat modelling, risk review, human factors issues, and hazard analysis is identified.

3.3.3 Interviews

The key for better understanding of perceptions, attitudes, meanings, suggestions, and opinions are interviews (qonita 2018). They comprise of a series of questions (sometimes open-ended) for the collection of facts and data (raw information).

Interview gives a deep insight knowledge about a domain by observing the expertise, attitudes and experiences of users as compared to *Focus Group* which are mainly meant to exchange viewpoints against a discussion or idea. Moreover, the reason for not choosing *Contextual Interview* is because not only 'how' but 'why' part behind a task (job) need to be studied. For example, during the interviews of 'Train Signaller' the major tasks performed are to be understood and human evaluation in-terms of tendency of humans to make error or mistakes is to be studied as well. For designing interviews following three principles are considered: explore for contextual inquiry, generate a participatory plan design, and evaluate for usability and human factors.

During this PhD, semi-structured interviews with relevant stakeholders including human factors experts, safety consultants, and ex-rail signallers are conducted. The selection criteria behind these stakeholders is dependent on scope of research such as a human factors expert for understanding TA, use-case specifications, human error and performance, a safety consultant for security aspects by identifying potential hazards leading to risk analysis, and ex-rail signaller for performing TA by giving details about working hours, task routine, major and sub-tasks performed along with cognitive attributes responsible for impacting performance.

These semi-structured interviews comprised of open-ended questions. There is no mandatory list of questions, but the intent is to conduct an inquisition of knowledge by initiating on open discussion.

3.3.4 Case Study Research

Several disciplines of science support the case study research methodology for determining the empirical truths using qualitative approaches as shown in Fig. 3.5 (Harrison et al. 2017). As this methodology is helpful towards drawing conclusions about this multi-

disciplinary research. Therefore, the safe, secure and usable design framework is validated using three case studies as mentioned in Table. 3.1.

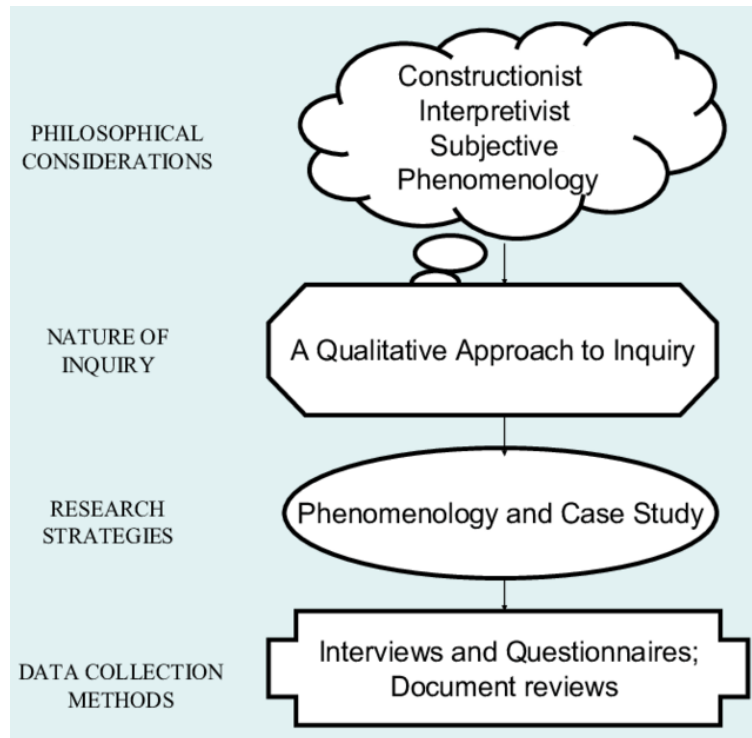


Figure 3.5: Qualitative Case Study Research Plan (Harrison et al. 2017)

First, a real-life case study of '*Polish Tram Incident*' is conducted, where the interdependencies between safety, security and human factors engineering are present. This case study is used to validate the linkages between different elements of safety, security and human factors like asset, vulnerability, threat, risk, task, goal, obstacle, safety hazard, and human error. Along with security expertise, the safety and human factors stakeholders from rail are involved throughout the process for feedback and support.

Table 3.1: Case Study Research for PhD

| Case Study Title | Synopsis | Chapter in Thesis |
|--------------------------------------|---|--------------------------|
| Polish Tram Incident | Validate the Integration of Security with Safety and Human Factors Engineering | 4 |
| A Day in the Life of ERTMS Signaller | Process-techniques and Tool-support Options to Validate Design Framework | 5 |
| Cambrian Incident Report | Design Framework is an Exemplar for Resolving Safe, Secure and Usable Design Issues in Rail | 6 |

Second, the results from first case study are used to inform TA using use-case specifications format by prototyping the role of *ERTMS - Signaller*, which provides human factors experts a chance to work in collaboration with safety and security design experts. The approach is derived from IRIS framework and CAIRIS tool-support. The aim is to

recognise different levels of human failures in the form of recognition of errors or mistakes, using CTA, HTA, and HFACS framework.

In the final case study, with the support of representative rail stakeholders from Ricardo Rail is used to implement STPA on case study of '*Cambrian Railway Incident*'. This case study helped design framework to act as an exemplar which is to serve as a tool guide for human factors, safety and security experts to deal with human factors issues, associated safety hazards, and potential security risks.

3.4 Summary

In this chapter, the research methods as stated in Section 3.3 are adopted to achieve research goal and validate the research questions as mentioned in Sections 1.2 and 1.3, respectively. First, the integration between safety, security and human factors engineering is achieved. Second, this integration is used to include process-techniques and tool-support for safe, secure and usable design framework. Finally, the proposed design framework is validated to ensure it acts as an exemplar for resolving safe, secure and usable design issues in rail infrastructure.

Chapter 4

Safe, Secure and Usable Design Framework

In this chapter, a design framework is devised based on process-techniques and tool-support from safety, security and human factors engineering. This design framework is based on integration of concepts between engineering domains as motivated in Chapter 2. This integration builds the foundation for resolving safe, secure and usable design solutions in rail infrastructure. Also, this aims to accomplish research goal which is to provide a single platform for design analysis to safety, security and human factors experts.

A preliminary evaluation of the security-by-design process-techniques for design framework with tool-support of CAIRIS is conducted. Using the research methodology mentioned in Section 3.3, the validation is carried out by applying it to a real-life case study of 'Polish Tram Incident'. The design framework is applied in three phases, where security elements like asset, role, attacker personas, vulnerability, threat, risk, task, and goal-obstacle are identified using IRIS framework, leading to potential safety hazards, and human error sources using HFACS framework.

4.1 Integration of Concepts

The security risks are due to hidden vulnerabilities within system leading to threats and cyber-attack possibilities (Brostoff and Sasse 2001). It has also been observed that the security mishaps in the systems are sometimes due to human failures within an environment (Jonsson and Olovsson 1998). These human failures are as a result of errors, mistakes or lapses made by humans, which are the determining factors for human performance as well (Felice and Petrillo 2011). Since safety engineers, focus on identifying all potential hazards as a result of security risks and human failures (Brostoff and Sasse 2001). Thus, security risks, human failures and safety hazards, each need to be considered during design of any critical infrastructure.

By design, those specifications of the system are meant which are responsible for determining the process required for product development. The design specifications for any system can be improved by minimising the chances of human errors or mistakes. Usually, these human errors, mistakes and lapses etc are classified as the determining factors for human failures, which can be identified using HFACS framework (Wiegmann and Shappell 2003).

The use of security-by-design approaches for risk analysis is achieved by the secure and usable IRIS framework using CAIRIS platform (Faily 2018). This meta-model with six views of system is responsible for integrating design concepts within security and usability (human factors). Also, CAIRIS platform helps to visualise different concepts by connecting them together. For example, during risk analysis the associated threats and vulnerabilities are identified along with assets in jeopardy. The potential attackers leading to these risks are assigned roles and their personas help to model potential goals, responsibilities, and tasks. Hence, helping to recognise threat and risk modelling elements in the form of attacker profile, vulnerability and threat identification.

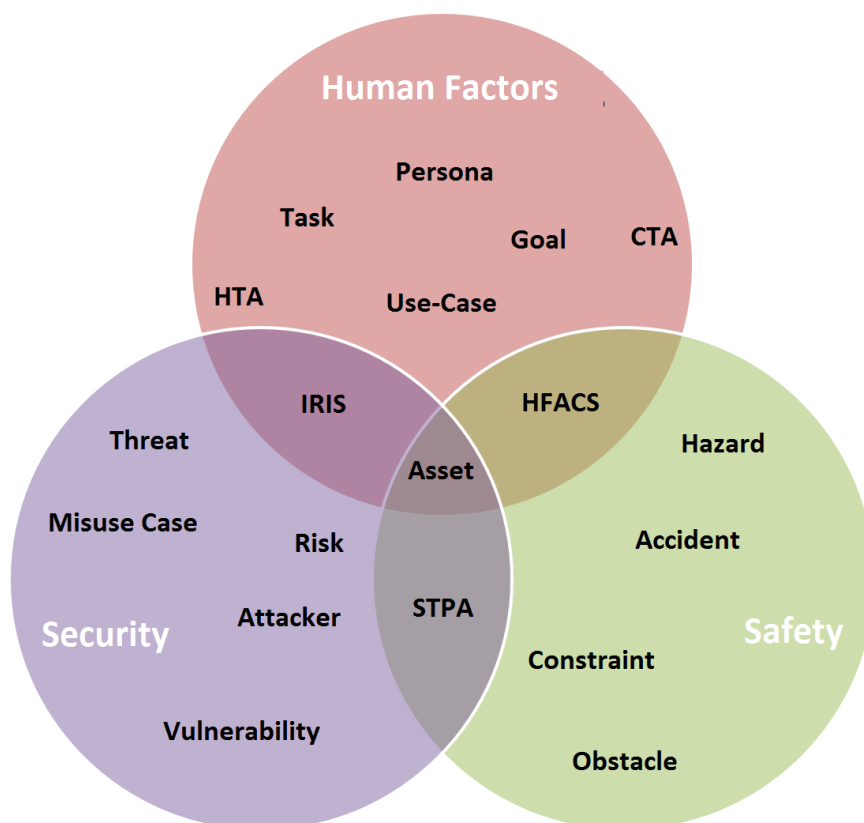


Figure 4.1: Integration of Concepts Between Safety, Security and Human Factors

Meanwhile, the human factors experts are using TA as one of the many approaches for making design improvement decisions based on human performance (Affairs 2013). Usually, TA is conducted on elicited tasks using User Experience (UX) design techniques

such as personas. The personas are based on roles within system and based on narrative scenarios. By proposing a use-case specifications template based TA as a combination of both CTA and HTA, the associated human failures (active and latent) can be identified and classified using HFACS framework.

The analysis from security-by-design approaches and human factors engineering contribute towards the identification of potential safety hazards. These safety hazards are then used to conduct STPA, for identifying control actions and causal factors behind accidents for improving system design. Also, STPA has the tendency to feed into vulnerability, threat and risk analysis for security.

Therefore, the research goal of this PhD thesis which is to propose a design framework based on best process-techniques and tool-support from safety, security and human factors engineering is achieved as shown in Fig. 4.1. The safe, secure and usable design framework comprises of three major sections:

1. Security-by-Design Approaches (IRIS and CAIRIS)
2. Safety Analysis (Hazard Identification)
3. Human Factors Engineering Techniques (HFACS Framework)

4.2 Security-by-Design Approaches

From security-by-design, a process-technique is devised where secure and usable IRIS framework is implemented to identify associated security, safety and human factors design issues. This process-technique is tool-supported using open-source CAIRIS platform, and brings forward design framework for safe, secure and usable design solutions by integrating concepts together.

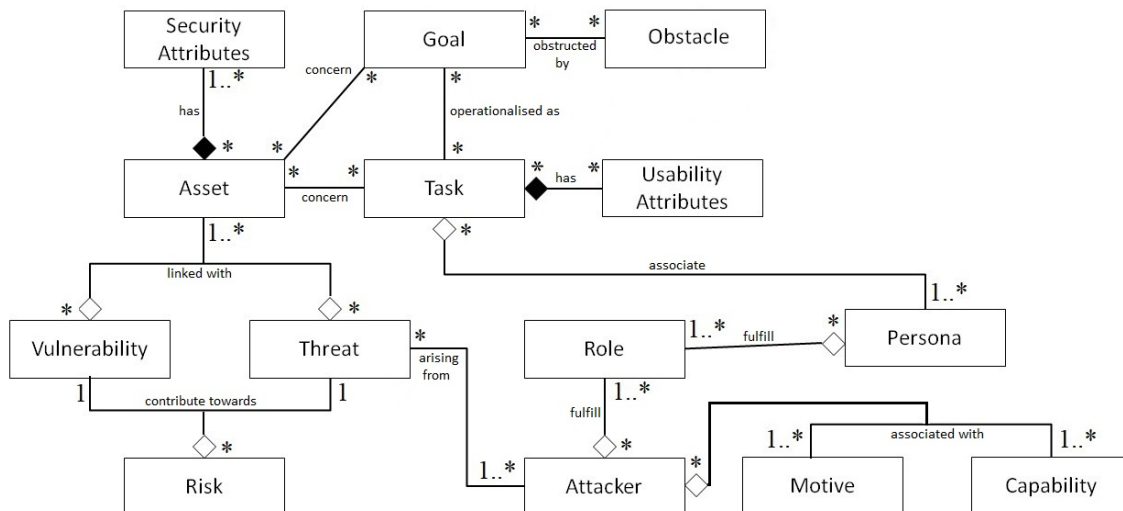







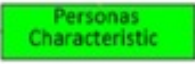


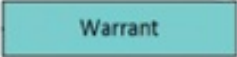
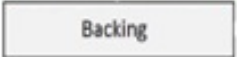





Figure 4.2: Security by Design Approach Consisting of IRIS Framework

| <i>Graphical Notation</i> | <i>Name</i> | <i>Description</i> |
|---|------------------------------------|--|
|  | Asset | An asset is a tangible object of value to stakeholder. |
|  | Asset with Security Property(ies) | An asset with security properties: integrity and availability. |
|  | Security Property: Confidentiality | The security property which aims to ensure that information is not available to unauthorised person. |
|  | Security Property: Integrity | The security property which aims to ensure the fulfilment of assets. |
|  | Security Property: Availability | The security property which aims to ensure the accessibility of assets (information or system etc). |
|  | Role | A role is a representation of human element (behaviour and act) within a system. The roles are defined based on stakeholder owning resources within system. |
|  | Personas | Personas describe the archetypical behaviour of system actors. |
|  | Personas Characteristic | Personas characteristics support the validity of personas and is categorised into: activity, attitude, aptitude, motivation, skill etc. |
|  | Modal Qualifier | A modal qualifier is used to describe the confidence in the validity of the personas characteristic. |
|  | Ground | Ground is evidence of the persona's validity. |
|  | Warrant | Warrant act as inference rule connecting the grounds to the personas characteristic. |
|  | Backing | Backing are document references acting as arguments known as <i>factoids</i> . |
|  | Task | A task represents the work (job) to be accomplished by personas. They have usability attributes and different values are represented by varying shades of colour blue. |
|  | Goal | A goal is a requirement to be fulfilled by system or user. |
|  | Responsible Refinement | A refinement type which acts as obligation between sub-goals. |



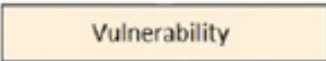







| | | |
|---|------------------------------------|--|
|  | And Refinement | Conjunction (union) of sub-goals. |
|  | Obstacle | Obstacle is an obstruction to goal. |
|  | Vulnerability | A vulnerability is a system weakness usually hidden beneath a system. |
|  | Threat with Security Property(ies) | A threat with security properties: confidentiality, integrity, and availability. |
|  | Risk | A risk is defined as the likelihood and severity of an incident. During risk analysis, the darker shades represent risks with high impact level depending on risk scoring. |
|  | Level of Human Failure: High | A use-case with <i>high</i> level of human failure. |
|  | Level of Human Failure: Medium | A use-case with <i>medium</i> level of human failure. |
|  | Level of Human Failure: Low | A use-case with low level of human failure. |
|  | Process Model | A process model highlights the design-level issues leading to accident scenarios because of hazard. |
|  | Control Algorithm | A control algorithm is an effective way of modelling control structures consisting of components (process) and controllers (feedback). |

Figure 4.3: CAIRIS Graphical Notation

4.2.1 IRIS and CAIRIS

The security and usability design concepts are better understood as suggested by IRIS process framework (Faily 2018). It is complemented by the CAIRIS platform, which acts as an exemplar for tool-support to manage and analyse design data collected when applying an IRIS process.

Using IRIS framework the security elements related to vulnerabilities and threats as they contribute towards potential risks are better understood (Faily and Flechais 2011). Another aspect is to view it from the attacker's perspective, where the intent contribute towards risk analysis. The attacker's intent using personas narrative helps to identify the tasks and goals they carry out, which when exploited helps determine human factors issues (active and latent failures). These security risks and human error sources might contribute towards potential hazards for safety analysis. Therefore, IRIS provides a foundation for integrating security, safety and human factors.

The IRIS framework and CAIRIS tool-support, which leverages security and usability

engineering approaches, are used to better understand the safety implications of any critical infrastructure under design. The proposed process-technique takes input from security and human factors engineers, as well as from relevant field stakeholders with safety expertise.

During this PhD, the following CAIRIS notations and graphical symbols are used as shown in Fig. 4.3.

4.2.2 Asset Modelling and their Associations

The process-technique begins with a security analysis of the system under observation and its environment by identifying the possible assets (Gollmann 2007). These assets and their relationships are modelled using UML class diagrams as shown in Fig. 4.4. Each asset is defined in a particular environment, and categorised by asset types. The asset types comprise of people, information, system, hardware, and software. The security attributes for assets like confidentiality, integrity, availability are defined and values (Low, Medium, High) are assigned, based on priorities defined by the stakeholders involved in design. The values are assigned according to priority. For example, an asset *Railway Staff* has confidentiality (of information) set as 'Low', integrity (ensure job) set as 'High', and availability (during job) set as 'High' too.

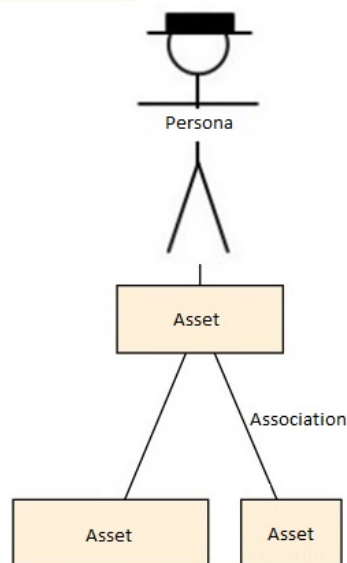


Figure 4.4: Asset Model Using UML

4.2.3 Role and Attacker Personas

The roles are defined based on stakeholder owning resources within system. The roles are further used to identify specific personas describing the archetypical behaviour of system actors.

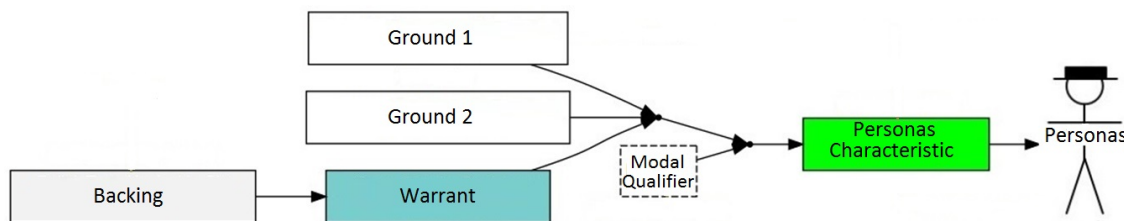


Figure 4.5: Argumentation Model for Personas Characteristic

Attacker personas are created by following the approach described in (Atzeni et al. 2011); 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 from Trello into CAIRIS as personas characteristic argumentation models. These argumentation models are based on Toulmin’s model of argumentation as shown in Fig. 4.5, 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.

4.2.4 Vulnerability Identification and Threat Modelling

The vulnerabilities are weaknesses of the system, which, if exploited, leads to a security breach (Faily 2018). 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 (confidentiality, integrity, availability) based on the goals of attacker.

4.2.5 Risk Analysis

Vulnerabilities and threats contribute to the identification of potential risks (Faily 2018). Using risk analysis, the likelihood and severity of an incident is determined based on the ability of an attacker, and the value of asset that need to be protected. At this stage, the evaluations are made based on existing security controls. 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 for further analysis as shown in Fig. 4.6.

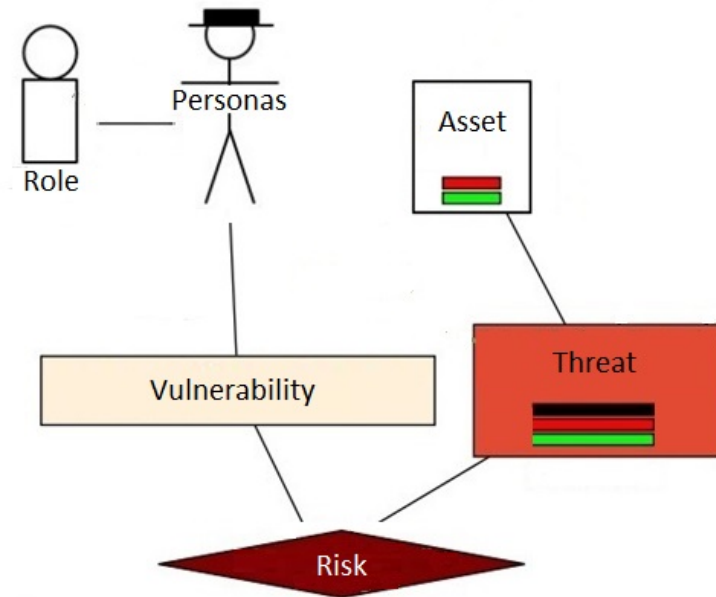


Figure 4.6: Threat and Risk Analysis Modelling in CAIRIS

4.2.6 Task and Goal-Obstacle Modelling

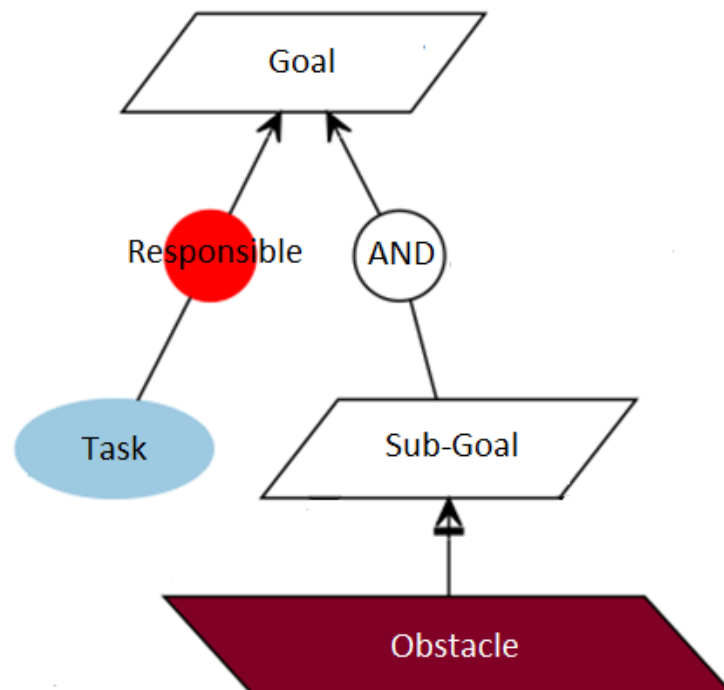


Figure 4.7: Task and Goal-Obstacle Modelling in CAIRIS

Based on asset modelling and risk analysis, the concerned tasks and goals are elicited. The concerned tasks are defined as the specific jobs performed by users which are identified using roles and personas. Goals are efficient fulfilment of these jobs. These form the basis of system and user level goals. Tasks and goals are identified from the attacker's perspective (personas) and form the basis for *obstacles* that model obstructions to system goals as shown in Fig. 4.7. Goal and task models can help the security engineers to better understand the system threat model. The approach used by IRIS and CAIRIS for modelling both goals and obstacles is based on the KAOS goal modelling language.

4.3 Identification of Safety Hazard

The risk model generated by CAIRIS determines the safety hazards. At this stage, the keyword *risk* refers to analysis (information security) as a result of threats and vulnerabilities (Section 4.2.5). It shows the linkage between the assets (to be protected) with their associated security attributes, the vulnerabilities (system weaknesses to be exploited), the emergent threats (based on attacker's capabilities and motives) and the possible risks (threat scenarios). 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.

Also, the human failures identified using HFACS framework as a result of TA, contributes towards the identification of potential hazards. The hazards may be based on human and system interactions, especially human errors or mistakes (Mindermann et al. 2017). Consequently, these hazards are classified as the major derivatives behind catastrophic accidents.

4.4 HFACS Framework

The HFACS framework presents four levels of human failures (error sources), for all categories of Reasons's error taxonomy based on cognitive, behavioural, personal, and organisational factors (Wiegmann and Shappell 2003).

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 framework. The value with the closest possible explanation for human error is labelled as the desired human factors issue.

Human Factors Integration (HFI) is an alternative systematic process that has been considered. HFI is a process for identifying, tracking and resolving human-related considerations ensuring a balanced development of both technologies and human aspects.

However, HFACS was better suited for this research due to the unsafe acts component i.e. handling of errors (decision, skill-based, perceptual) and violations (routine, exceptional).

4.5 Case Study - Polish Tram incident

The aim of the case study is to validate the integration of security with safety and human factors engineering. The safe, secure and usable design framework is applied for empirical evaluation using case study analysis. This analysis aims to evaluate and validate the process-technique and tool-support behind design framework.

The 'Polish Tram Incident' is an example where the security breach into a system was responsible for compromising the safety of people in the form of passengers of tram. 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 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 (RIS 2008)".

4.5.1 Overview

From the perspective of security engineering, this incident can be dissected to reveal the associated safety implications. Generally, there is a lack of incidents in rail with safety implications because the safety and human factors are already prioritised in rail industry. But with the technological advancements especially by considering open networks there is a lack of consideration towards security. Hence, this incident of 'Polish Tram' took place.

For this purpose a high-level architectural overview of the incident is presented as shown in Fig. 4.8. A case study based on light rail is chosen rather than heavy rail for three reasons. First, different categories of actors (passenger, staff, attacker) interacted with system (tram service) for their own particular purpose as is visible from the architectural overview. As such, this incident intersected safety, security, and human factors engineering, providing an opportunity to examine it from security engineering perspective for possible system and environment flaws which remained unexplored. Second, the consequences of this incident or accident can be easily generalised using online information and data. For example, 4 trams were derailed and 12 people were injured. This stresses that security of assets (trams and network) and safety of people were compromised. Finally, because the public use is comparatively more frequent, light rail has an increased

likelihood of hazards due to possible errors and mistakes leading to catastrophic consequences.

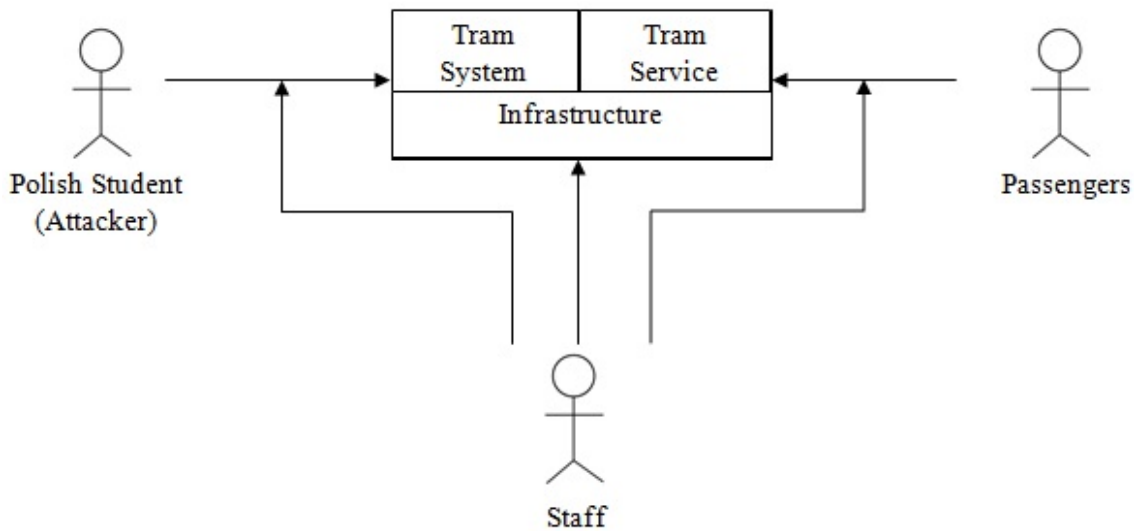


Figure 4.8: High-level Architectural Overview for Polish Tram System

4.5.2 Attacker Perspective

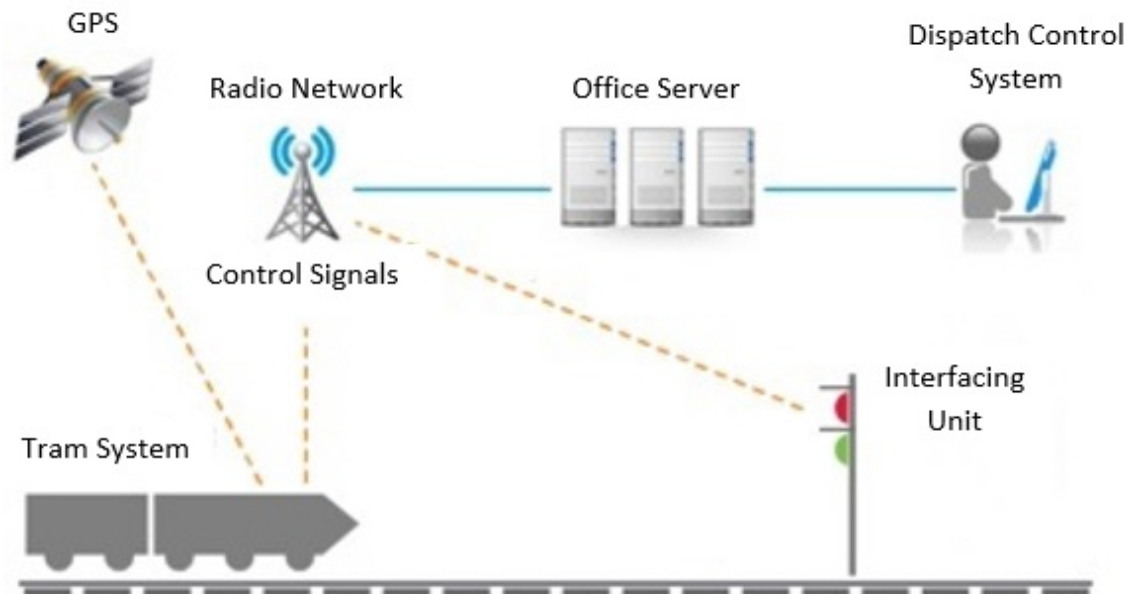


Figure 4.9: Working Infrastructure of Polish Tram System

The attacker *Adam Dabrowski* was a 14 year old boy and a brilliant student of mechanics, this initiates the ground for motivation behind this attack. Another point of consideration is the attacker's intent, where he did not intend to harm people in any way. Usually,

the attacker was found exploring the city's tram system. Out of passion and curiosity, the attacker started using the tram depots for gathering desired information along with equipment required for building an Infra-red (IR) device. The attacker lacked the financial resources and was solely relying on the open-source information from public libraries like the Internet.

In 2008, the Polish Tram system was operating on 1970s switching system. The radio communication system was working on Advanced Train Control Protocol System (ATCPS) communication protocol for changing the track switches and controlling locomotive movement (Papa and Shenoj 2008) as shown in Fig. 4.9. From an attacker's perspective, the attacker's first priority was to dissect the working infrastructure of Polish Tram and look for possible vulnerabilities within working system.

As a result, the attacker was able to build a universal IR remote control. The simple concept of *replay attacks* was used to get access of the signalling system. The attacker used the remote control to record signals sent in one place to a set of points, and was able to replay the signals back to get similar results in another place.

Many of the universal remote controls have a learning mode, where you hold them in front of the original remote control, and it captures the signal. There is no requirement for learning complex encoding types, it just samples the signal at a sufficiently high rate to be able to replay it bit-by-bit. However, the only exception was increasing the power of Light-emitting Diode (LED) current limiting resistor, by adding a transistor along with more LEDs.

4.5.3 IR Remote Control

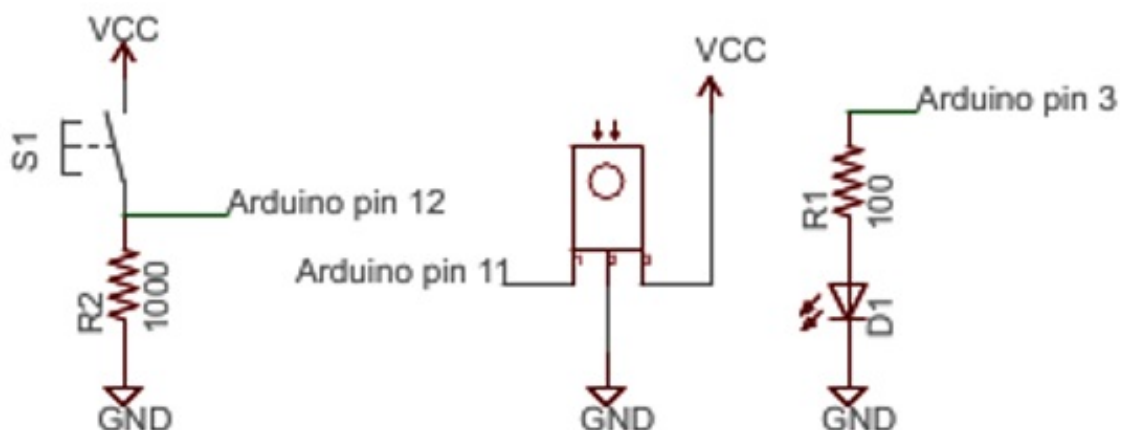


Figure 4.10: Hardware Setup for IR Remote Control (Ard 2009)

A simple IR universal remote control can be easily modified to record an IR code and re-transmit it as required. For this purpose, the hardware required is 9V battery pack, the arduino board, and the proto board with (top-to-bottom) the IR LED, IR receiver, and

push-button. The circuitry is simple, where an IR sensor module is connected to pin 11 to record the code, an IR LED is connected to pin 3 to transmit the code, and a control button is connected to pin 12 (Ard 2009) as shown in Fig. 4.10.

The code is freely available over the Internet and can be easily downloaded. In order to use the universal remote, simply point the remote control at the IR module as a source of transmission and press a button on the remote control for recording. Afterwards, press arduino button for retransmission of the code.

4.5.4 Cyber Attack

The cyber attack as shown in Fig. 4.11 was conducted by switching movements (directions) of tram from the middle point i.e. the front part moved in one direction and the rear part in the other. This splitting of switch caused tram derailment. According to online documentation, the points were safeguarded against the pressure sensors under the tracks. Therefore, the attacker tried multiple times until he found a fault track point. The points at Lodz were faulty against the system safety preventing the points from moving.

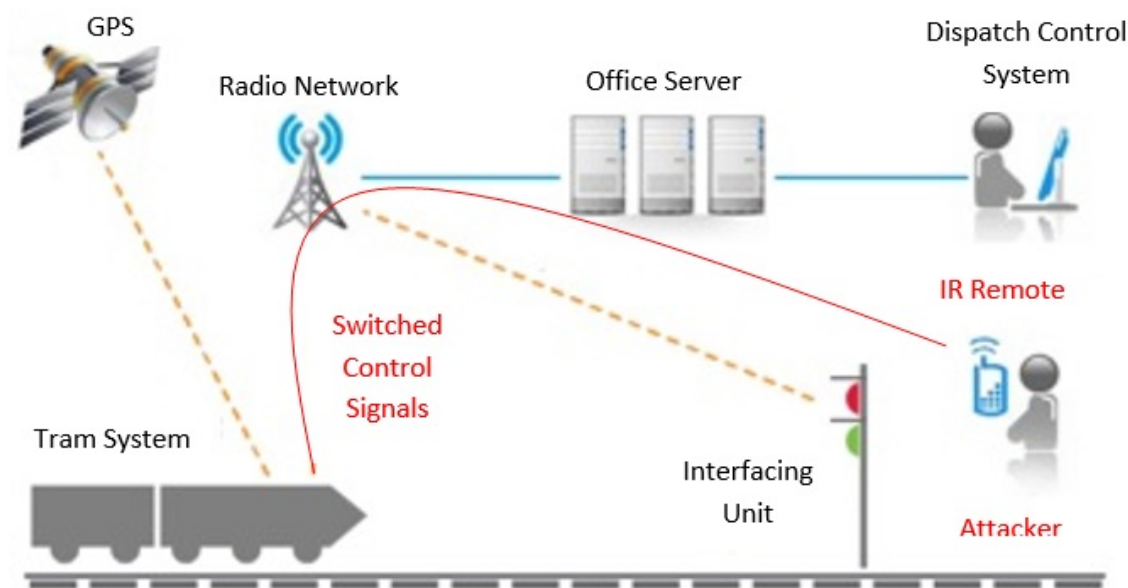


Figure 4.11: Cyber Attack for Polish Tram System

The occurrence of splitting of switch was possible due to lack of safety case, where tram occupying the switch should not have moved. In this safety case, the security failing would have been the assumption that the IR transmission component failure was a non-malicious event. Therefore, based on mean time between failure data rather than forced by a malicious attacker with higher probability (Papa and Shenoj 2008).

An accurate safety case would have been to incorporate pressure sensors within track points. where the trams send electricity over the rails to make a point switch. When the

tram enters the segment just before the point, the point is reset to the straight position by a pressure sensor (Papa and Shenoï 2008). If an electric pulse is sent over the rail, the point switches to the diverging position with a very audible click to confirm that.

4.6 IRIS and CAIRIS

An open source intelligence was gathered as an input to this methodology of IRIS framework. This was based on several online articles written about the particular Polish Tram Incident. The methodology was supplemented by publicly available data; this was used to understand the system architecture, application levels, operating modes, signalling principles and control. During this research, feedback on the emerging CAIRIS model was obtained from safety and human factors experts at Ricardo Rail, who were representative of the rail stakeholder.

4.6.1 Asset Modelling and their Associations

In rail, integrity is typically the most important security attribute of the system followed by availability and confidentiality (Bloomfield et al. 2016). The assets were identified based on online-data sources about the incident of Polish Tram and their associations were defined keeping in mind the rail infrastructure. Two working environments were defined: *Morning* and *Night shift*. The *Morning Shift* is from 0600-1800 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*, which is from 1800-2400. 51 assets were identified, based on types of software, hardware, information and people as shown in Fig. 4.12.

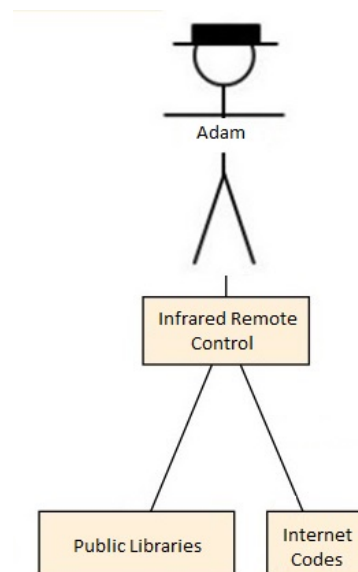


Figure 4.12: Asset Model Using UML

The assets and their associations helped define the roles and personas (*attacker, driver and signaller*). They also helped to identify the tasks and goals that need to be fulfilled. 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.

4.6.2 Role and Attacker Personas

The analysis about the rail infrastructure lead to the recognition of 11 major roles as shown in Table. 4.2. The most notable was the role of *Attacker*.

Based on online articles and incident records as shown in Table. 4.1, it is concluded that the attacker did not wish to intentionally cause harm. Instead, attacks were exploratory in nature with no consideration given to the consequences. Attacker also lacked the funding and adequate resources to conduct the attack. Curiosity and passion were identified as the major motivations and attacker was only equipped with basic knowledge about the information and railway sector. The role of attacker further motivated to understand the intent and capability behind the cyber attack with the help of personas.

Table 4.1: Online Articles Used as Data Source for Building Attacker Personas

| Ser. | Article Title | Author | Publisher |
|-------------|---|-------------------|---|
| 1. | Hacking Polish Trams | Bruce Schneier | Schneier on Security Article |
| 2. | Polish Teen Derails Tram After Hacking Train Network | John Leyden | The Register Article |
| 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 Article |
| 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 Article |

Table 4.2: Major Roles in Rail Infrastructure

| <i>Rail Sector</i> | <i>Role</i> | <i>Description</i> |
|---------------------------|-------------------------|---|
| Engineer | Locomotive Engineer | The 'Locomotive Engineer' is a stakeholder responsible for managing the trams/ locomotives involved. |
| Railway System | Train Driver | The 'Train Driver' is responsible for driving the trams. |
| | Railway Manager | All the management issues related to railways are solved by 'Railway Manager'. |
| | Railway Security | The security of the railway station and their affect on passengers and staff present are looked after by the security guards. |
| | Railway Ticketing Staff | The ticketing system is responsible for issuing tickets to passengers who desire to use the tram system. |
| | Railway Passenger | The general public who is using the railway system to get from one place to another. |
| Operations | Railway Dispatcher | The stakeholder who is responsible for managing the railway traffic signals. |
| | Train Signaller | The stakeholder who is going to monitor the railway switches and junctions. |
| Maintenance | Track Inspector | The stakeholder who is going to look after the maintenance requirements of tracks on which trams are going to run. |
| | Power Supply Manager | The stakeholder who is going to make sure that uninterrupted power is supplied during railway operations. |
| Attacker | Cyber-Attacker | A 14 year old boy, a Polish student, hacked into the tram system which enabled him to track points. Four trams were derailed. 12 people were injured. |

An attacker personas *Adam* was created 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. These characteristics help understand threats the infrastructure afforded to *Adam*.

For example, Fig. 4.13 illustrates the argumentation model underpinning the persona characteristic *Working Knowledge about Railways*. *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, 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* was identified.

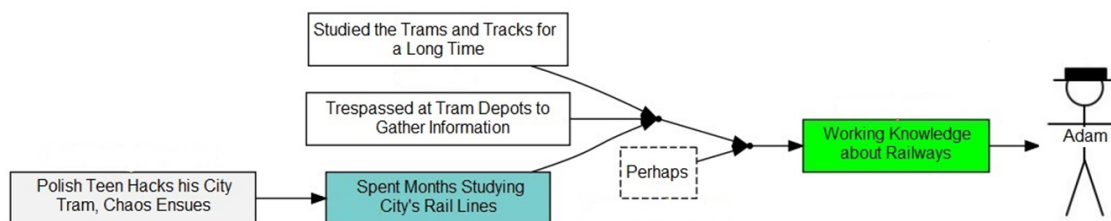


Figure 4.13: Argumentation Model for Personas Characteristic

4.6.3 Vulnerability Identification and Threat Modelling

The ulterior motive of *Adam* (*get into tram network*) by compromising the security of assets (*switches and rail junctions*) was achieved by leveraging a system weakness (*faulty track points*). By exploring the 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.

3 threats were also identified: *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 interpretation of *Adam's* ability to exploit *Faulty Track Points*.

4.6.4 Risk Analysis

The emergent risks based on threat model also formed the basis for misuse cases (threat scenarios) stating how *Adam* was going to interact with an environment. Within an environment of *Morning Shift*, 4 risks and misuse cases were defined using vulnerabilities and threats. These form the basis of the risk model, the results of which are illustrated

in Fig. 4.14. 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. In *Morning Shift*, due to the presence of a large passenger numbers, the risk of occurrence of *Train Derailment* has more impact on passenger safety. It could be life threatening for staff and passengers both on and near trains, which eventually occurs due to a negligence on the part of security engineers to correct the known problem of faulty track points in time.

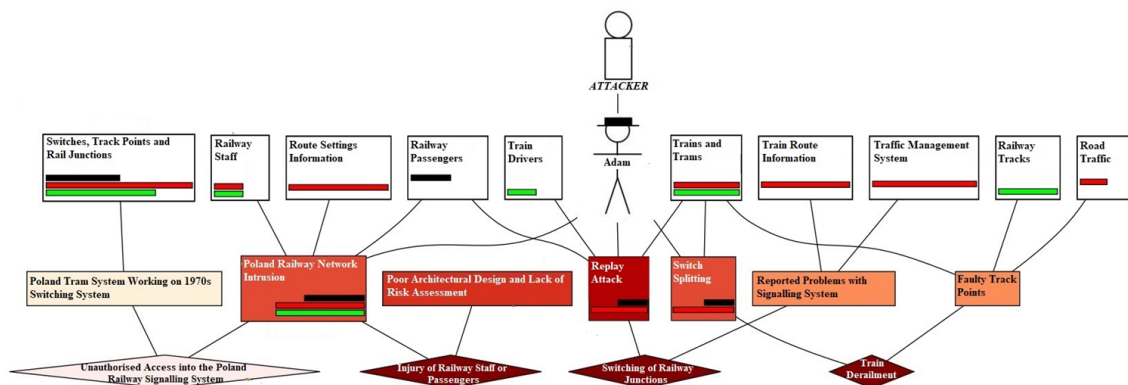


Figure 4.14: Risk Modelling in CAIRIS

4.6.5 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, *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 as shown in Fig. 4.15.

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.

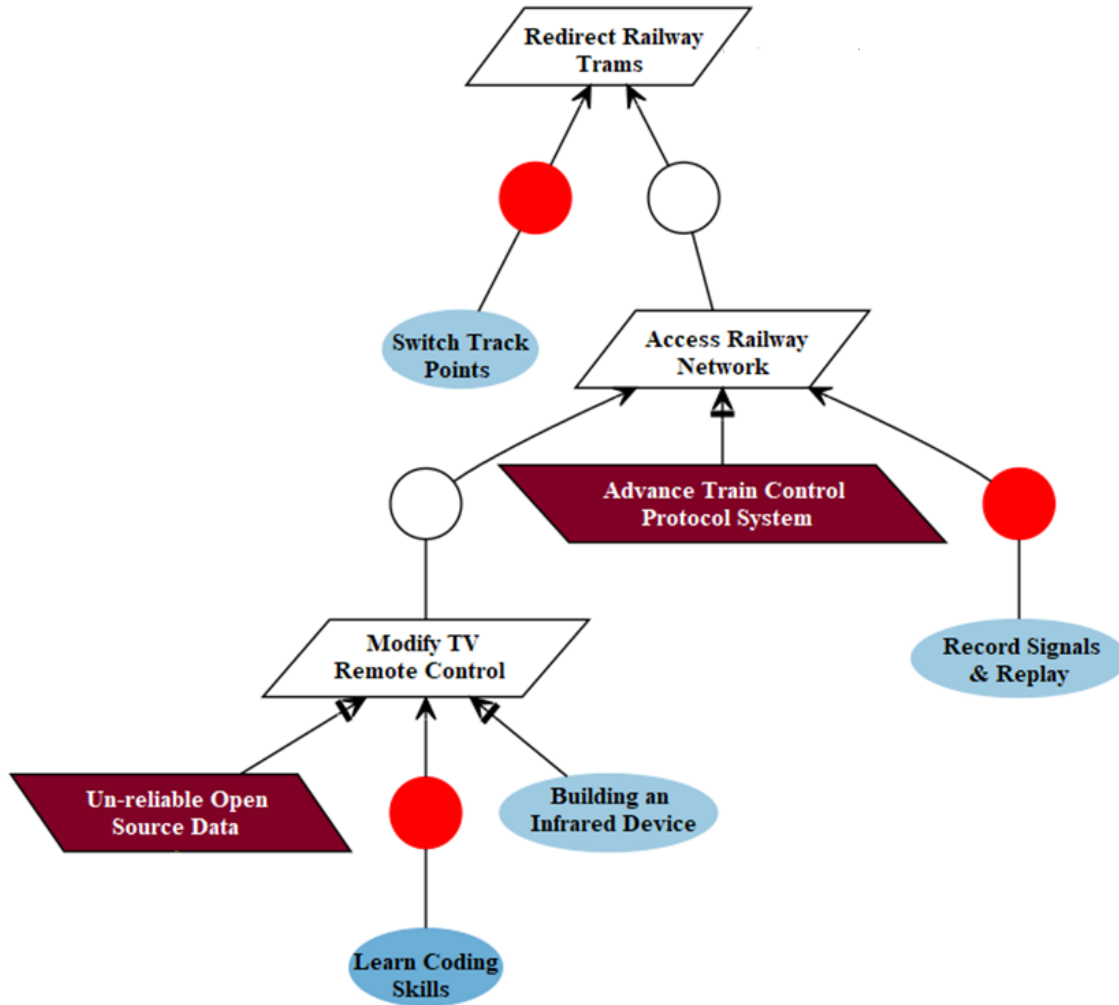


Figure 4.15: Task and Goal-Obstacle Modelling in CAIRIS

4.7 Identification of Safety Hazard

For explanation purposes, 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. Similarly, the occurrence of the risk of *Unauthorised Access into Poland Railway Signalling System* might lead to *Disruption of Train Services*.

Table. 4.3 represents the identification of potential safety hazards from risk modelling elements (vulnerabilities, threats, risks) based on the Risk Assessment Log presented by Randstad Rail (Ran 2014). The documentation of Randstad Rail, includes the activities and tasks in railway environment 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 about problems. As safety engineers, are concerned with all those aspects of system design where unintentional actions by benevolent actors may compromise human safety (Young and Leveson 2014).

4.8 Human Error - HFACS

The human error is considered the biggest source behind active failures and, by identifying appropriate errors, the relevant human factors issues can be resolved, for instance by the implementation of user training and standardisation, etc.

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. Here, the timely evaluation of technological environment using checklists and task factors can minimise the chances of risk occurrence.

Table. 4.3 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 to verify the system usability for risks, by the satisfaction of user goals depending on certain procedures, competencies, permissions and TNA to achieve those goals and complete defined tasks.

Table 4.3: Human Factors Issues based on HFACS

| <i>Vulnerabilities</i> | <i>Threats</i> | <i>Associated Risks</i> | <i>Safety Hazards</i> | <i>Human Factors</i> |
|---|---|---|--|----------------------------------|
| 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 |

4.9 Discussion

The persona of *Adam* is built keeping in mind the security and usability aspects using user-centered design approach. This gives an idea about the possible thinking of an

attacker; by providing security engineers a better chance to identify the system vulnerabilities which can lead to threats using an attacker-centric view of system. In this way, using threat modelling features of CAIRIS the anticipation of possible risks are made.

From 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*.

Sometimes the emergent threats have the tendency to risk the safety of environment along with security of system. Using asset modelling and their associations, attacker personas, task and goal-obstacle modelling, this overlap between safety and security is well understood.

Adam was able to cause a security breach through which he intruded the railway network and switched tram directions. As a result, twelve people got injured. No deaths occurred, due to timely action by the 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 major safety hazards faced due to security breach. *Adam* was also charged for endangering public safety, which further confirms the potential safety hazard.

Here, STPA can be linked with IRIS framework to classify the identified hazards in more detail and determine the possible control actions for them. The identified hazards may be as a result of human errors or mistakes, for which HFACS framework can be used. This motivates the integration of safety concepts into IRIS framework and CAIRIS tool-support.

The assets, roles, personas, vulnerabilities, threats, risks and safety hazards modelling helped to visualise the possible tasks scenarios. Using these task scenarios the usability parameters namely, task duration, frequency, demands etc., of the users were determined using CAIRIS. Moreover, the tasks further helped to identify the system level goals and user level goals. Thus assisting to better visualise the system, by presenting link between security and usability (human factors) in the form of goal-obstacle and responsibility modelling. Using CTA, the identified cognitive attributes responsible for affecting the task performance can further help to determine the human error sources using HFACS framework.

For example, *Adam learned the coding skills* from his class and internet before, *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.

Also, during Case Study I, the approach recognition and application was happening, simultaneously. This was due to complex nature of concepts which were explored while working on approach for case study. Here, along with literature survey the consultation from field experts during application of case study was helpful. Using their feedback and review a lot of issues were timely resolved such as recognition of concepts, selection of case study, and analysis of feedback. Similarly, dealing with a wide range of experts from safety, security, and human factors engineering during this research was another challenge. As each expert has a certain knowledge base based on difference of opinion. Here, the target was to bring them all together where each stakeholder's input and feedback was valuable and essential for research. Meanwhile, working on two research papers for publication and writing of *Major Review* for submission was also in progress.

4.10 Summary

In this chapter, a tool-support approach is presented as part of design framework, based on core concepts from IRIS framework and CAIRIS. The scientific novelty has been the methodology application to safety and human factors engineering in rail. A preliminary evaluation of this approach is carried out by applying it to a case study where inter-dependencies between safety, security, and human factors were present.

In doing so, three contributions have been made. First, the approach shows how asset modelling and their associations, allow assets to be prioritised by rail stakeholders. These assets are prioritised using different values of security attributes which are assigned by security stakeholders. The novel contribution has been the involvement of security stakeholders for asset valuation (estimation) and modelling. Second, this has been shown how building models of attackers contributes 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, process-technique using tool-support has shown how by 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.

Moreover, evaluation of this approach as part of design framework will be done on a project where the representative rail stakeholders will be more closely involved when considering the risks, roles, tasks, goals, requirements, dependencies and obstacles between the humans and systems. As next steps, a refined process-framework based on best practices from safety, security and human factors engineering are incorporated. 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 a more detailed in-sight into human factors, and subsequent security and safety concerns.

Chapter 5

Extension of Design Framework

In this chapter, a preliminary evaluation of the human factors process-techniques for extension of design framework with tool-support of CAIRIS is conducted. The validation is carried out by using Operational Concept of ERTMS specifications for the role of 'Signaller'. The design framework is extended by conducting TA as a combination of CTA and HTA using use-case specifications template. This case study is concluded by specifying the process-technique requirements for tool-supporting design framework.

5.1 Human Factors Engineering Techniques

The achievement of security for safety critical systems, in itself is a human factors driven concept. As human element in the form of interactions with the system, within an environment of any critical-infrastructure cannot be ignored at any cost. There have been more than enough research evidence showing the dependency of human factors with safety and security engineering (Sections 2.2, 2.4, and 2.6). Thus, human factors engineering driven processes and techniques can be considered as a way forward for accomplishing better safety and security goals.

The process-technique take its lead from task modelling in IRIS, and begins by performing TA as a combination of CTA and HTA for identifying human failures. In this phase, an approach is devised based on UX design techniques (such as personas) for task elicitation and use-case specifications informed TA as shown in UML class diagram in Fig. 5.1.

Another aim of this research is to provide a comprehensive technological software tool, in order to assist safety, security and human factors engineering experts by making their job efficient and easy. For this purpose, CAIRIS has been extended to incorporate additional concepts of TA in the form of CTA and HTA. This implementation in CAIRIS is also explained in subsequent sections.

First, the personas narrative elaborates the task performed by a role, which helps

to identify tasks for TA. Second, TA is conducted using a use-case specification pre-defined format. Finally, for each use-case specification CTA and HTA is performed. CTA is conducted by scoring relevant cognitive reactions. This leads to identification of different levels of human failures with the use of *Algorithm 1*. During HTA, associations between use-cases are identified. After colour coding of the use-cases, graphical models are generated using *Algorithm 2*.

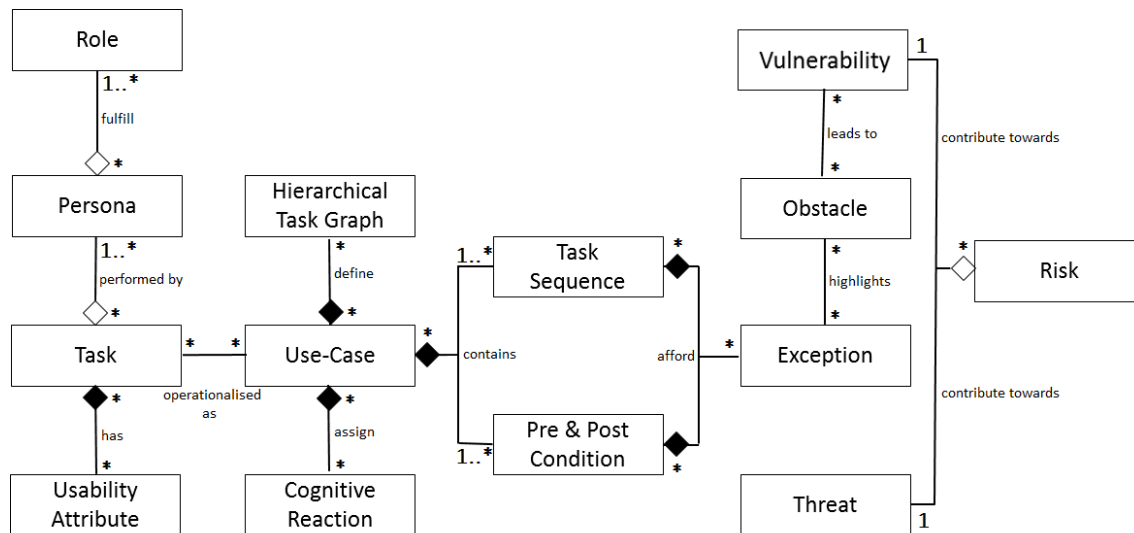


Figure 5.1: UML for Use-Case Specifications Informed Task Analysis

Consequently, these use-case models with specified level of human failures will help security and safety engineers better understand the associated risk modelling and safety analysis elements, like vulnerabilities, threats and potential hazards. Also, the use-cases with highest level of human failures will be categorised using HFACS framework to inform specific human error sources.

5.1.1 Personas for Task Elicitation

Personas are based on the Toulmin's Argumentation Models (Grounds, Warrants and Rebuttals), which aim at providing proper structure and assurance for qualitative data analysis (Atzeni et al. 2011). This approach is automated by using tool-support, such as CAIRIS as explained in Section 4.2.3. Using these argumentation models, the personas characteristics are identified, and scenario-based narrative is written (Faily 2018).

These personas narratives are used to elicit tasks for TA. After task elicitation, the relevant stakeholders are presented with organised information in rough tabular forms and their feedback is used to validate data for writing proper use-case specifications for conducting TA.

5.1.2 Use-Case Specifications Informed Task Analysis

TA is conducted for elicited tasks. For this purpose, use-case specifications are used as data gathering, representation and analysis tool. Use-cases allow both the user and functional characteristics of system to be presented, simultaneously.

| | | |
|--|---|--|
| Use Case Title | Name of use-case specification | |
| Abbreviated Title | Short name of use-case. | |
| Use Case ID | Identification number of use-case. | |
| Actors | A role/s played by user/s that interacts within system. | |
| Objective: Goal of interaction between user and system. | | |
| Pre-Conditions: State of system required before use-case. | | |
| Task Sequence: | Exceptions: | |
| 1. Normal sequence of steps involved within use-case. | An alternative flow or step involved within use-case. | |
| Post-Conditions: State of system required after use-case. | | |

Figure 5.2: Use-case Specification Template for Task Analysis

First, *Microsoft Excel* is used for preparing data spreadsheets, based on the anecdotal experience of most of the human factors engineers. A python script is developed for converting excel into Extensible Markup Language (XML) for importing use-cases into CAIRIS for further analysis and modelling of data. Second, the set of attributes are defined for the preparation of use-case specifications as shown in Fig. 5.2, including use-case title, abbreviated title, use-case id, actor/s, objective, pre and post condition/s, task sequence and exception/s. The choice for these attributes is based on two components of the system i.e., *user* and *function*. These selected attributes are required to simplify the complexity, by making it easier for stakeholders to read, understand and analyse use-cases. Finally, the use-case specifications are presented to human factors experts for validation through feedback. Afterwards, these use-case specifications are imported into CAIRIS.

Table 5.1: Cognitive Reactions and Performance Shaping Factors

| <i>Cognitive Reaction</i> | <i>Performance Shaping Factors</i> |
|----------------------------------|---|
| Vigilance | Tiredness, emotional stress, tension and fatigue. |
| Situation Awareness | Skill-set of an individual and HMI design. |
| Workload | Skills, HMI design, rules and guidelines. |
| Stress | HMI design, rules and guidelines. |
| Risk Awareness | Safety awareness, rules and guidelines. |

5.1.3 Cognitive Task Analysis

CTA is conducted for recognition of cognitive reactions, against each use-case. Previous research has shown that 5 cognitive reactions are found to have an influence on human performance based on Performance Shaping Factors (PSFs), such as tiredness, emotional tension, skills, Human Machine Interface (HMI) design, rules, guidelines, and safety awareness (Hammerl and Vanderhaegen 2009). Therefore, in this work 5 cognitive reactions are regarded as evaluators for human failures, namely: i) vigilance, ii) situation awareness, iii) workload, iv) stress, and v) risk awareness as described in Table. 5.1.

For each use-case values are assigned to these cognitive reactions such as *Low*, *Medium*, *High* or *None*, based on rationale. The stakeholders are given option for selecting one of these values against each use-case. For example, for a use-case of 'Combine Workstation' the vigilance is 'Low', situation awareness is 'High', workload is 'Medium', stress is 'Null', and risk awareness is 'High'. The stakeholder is free to choose any values as per priority and present a description as a rationale.

For this purpose, semi-structured interviews with relevant stakeholders are used for data analysis. These semi-structured interviews comprise of open-ended questions. There is no mandatory list of questions, but the intent is to conduct an inquisition of knowledge by initiating on open discussion. The stakeholders are presented with the proposed use-case specifications, where they are asked to select different values for cognitive reactions and document a rationale.

Algorithm 1: Level of Human Failure for each Use-Case

Data: u - the use-case specification

Result: l - the level of human failure for u

```

1 Function failurelevel(u) is
2    $sum = 0;$ 
3   for  $n \leftarrow 1$  to 5 do
4      $sum += cognitive\_reaction[n];$ 
5    $mean \leftarrow round(sum/5);$ 
6   if  $mean \leq 1$  then
7      $l \leftarrow Low;$ 
8     break;
9   if  $mean == 2$  then
10     $l \leftarrow Medium;$ 
11    break;
12  if  $mean == 3$  then
13     $l \leftarrow High;$ 
14    break;

```

Using values of cognitive reactions stored in database of CAIRIS, *Algorithm 1* is used for determining different levels of human failures. The algorithm takes each use-case as input and provides level of human failure for that specific use-case as output. For each use-case, *cognitive_reaction[n]* returns an array of 5 values of cognitive reactions where *n* ranges from 1 to 5. The values of *cognitive_reaction[n]* varies from *High*, *Medium*, *Low* or *Null*. Also, the values are associated with numbers such as, (0 for *Null*, 1 for *Low*, 2 for *Medium* and 3 for *High*). The mean (ranging from 0 to 3) of these cognitive reactions is calculated from *Sum* and rounded-off. Using mean different levels of human failures are determined. Mean is suitable as opposed to median which only points out the middle value while ignoring the individual value behind each cognitive reaction. Whereas, mode determines extreme values either too high or too low, eventually leaving mean as the best measure of central tendency. There are three levels of human failures against mean, 0 or 1 for *Low*, 2 for *Medium* and 3 for *High*, where *Low* being the use-case with less chances of human failure and *High* being the use-case with extreme chances of human failure.

5.1.4 Hierarchical Task Analysis

The task hierarchy is understood from task sequences as stated in use-case specifications. The high-level use-cases and tasks are divided into low-level use-cases and tasks, where each use-case is filled in with a particular colour depending on level of human failure assigned to it. The colour mapping is as follows: *dark blue*, *blue* and *light blue* for *High*, *Medium* and *Low* level of human failure, respectively. Using these colour codes, the different levels of human failures are better illustrated with HTA graphs using *Algorithm 2*. These different levels of human failures have the tendency to highlight use-cases and tasks, which require more attention by human factors, safety and security experts for design analysis.

The *Algorithm 2* takes no input instead its output is a set of quadruples i.e., (*h*, *h_fl*, *t*, *t_fl*) in which *h* is the head task name, *h_fl* is the head task failure level, *t* is the tail task name, and *t_fl* is the tail task failure level. The empty sets are defined for the quadruples *hta*, and task node/failure level pairs *visited* while enumerating the set (lines 2 & 3). The *buildTaskGraph* is a function that generates set of tuples from the CAIRIS model. Using this function, the algorithm entails by retrieving a set of tuples (*h,t*) in which *h* is the head task name and *t* is the tail task name. Each tuple in *buildTaskGraph* is enumerated, if *h* intersects with the first element in *visited* set then the task node/failure level from the set is retrieved (lines 6 & 7). Else, the *failurelevel* using *Algorithm 1* is calculated for the task node and *union* of task node/failure level with *visited* set is done (lines 9 & 10). The same thing is repeated for *t* (lines 12-17). Once we have tuples for *h* and *t* then quadruple is constructed by performing *union* with quadruple set *hta* (line 18). At the end, quadruple set is returned (line 20).

Algorithm 2: Build HTA Graph**Input** : None

Data: tg - set where each element is a tuple (h,t) in which h is the head task name and t is the tail task name, tt - tuple drawn from tg , $visited$ - set where each element is a tuple (t,fl) in which t is the task name, and fl is the task failure level, h_fl - tuple (h, fl) where h is the head task name and fl is the head task failure level, t_fl - tuple (t,fl) where t is the tail task name and fl is the tail task failure level.

Output: hta - set where each element is quadruple (h, h_fl, t, t_fl) in which h is the head task name, h_fl is the head task failure level, t is the tail task name, and t_fl is the tail task failure level.

```

1 Function buildHTAModel is
2    $hta \leftarrow \emptyset;$ 
3    $visited \leftarrow \emptyset;$ 
4    $tg \leftarrow \text{buildTaskGraph};$ 
5   while  $tt \leftarrow tg$  do
6     if  $tt[0] \in visited$  then
7        $(h, fl) \leftarrow visited\ tt[0];$ 
8     else
9        $(h, fl) \leftarrow failurelevel\ (tt[0]);$ 
10       $visited \leftarrow visited \cup (h, fl);$ 
11    end
12    if  $tt[1] \in visited$  then
13       $(t, fl) \leftarrow visited\ tt[1];$ 
14    else
15       $(t, fl) \leftarrow failurelevel\ (tt[1]);$ 
16       $visited \leftarrow visited \cup (t, fl);$ 
17    end
18     $hta \leftarrow hta \cup (h, h\_fl, t, t\_fl);$ 
19  end
20  return  $hta;$ 
21 end

```

5.2 Implementation in CAIRIS

During this PhD, a few implementation changes are made in CAIRIS. For demonstration of how TA approach using CTA and HTA is tool-supported in CAIRIS, the GitHub repository have been forked. The open-source CAIRIS installation guide is available at link:

<https://cairis.org>. The forked GitHub repository with implementation of *Algorithm 1* and *2* is available at link: <https://github.com/s5121191/cairis>. By installing this forked repository, the implementation details can be reviewed.

Using a high-level architectural overview of CAIRIS, the major components are CAIRIS Graphical User Interface (GUI), CAIRIS database, graphical model generation and view as shown in Fig. 5.3 (Faily 2018). The CAIRIS GUI provides an interaction medium to its user which is used to insert data into CAIRIS database. The scripts for graphical model generation are responsible for retrieving data and generating models.

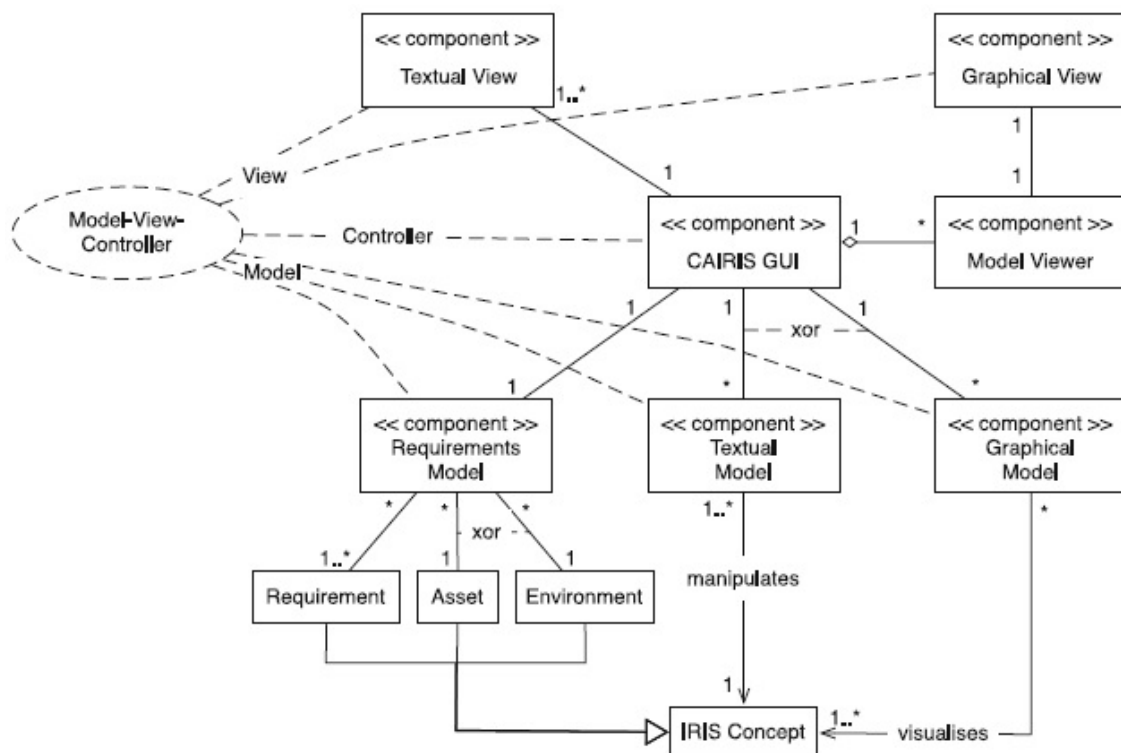


Figure 5.3: CAIRIS Architecture (Faily 2018)

For implementing the TA approach using CTA and HTA in CAIRIS by extending it, following steps are followed:

5.2.1 Development Environment

The development environment is set-up by installing open-source *VirtualBox*. The virtual machine consisting of *Ubuntu* desktop .iso image is created and started. Also, the GitHub pull request is initiated for forking CAIRIS repository. Using this repository, virtual machine is used to clone latest version of CAIRIS server.

5.2.2 Database Tables and Procedures

There is an independent MySQL database maintained behind each CAIRIS concept and model. Therefore, in order to support cognitive attributes and their values for each use-case specification new tables and database procedures are created. These database changes are defined for adding, updating, deleting and retrieving cognitive attributes and use-case associations. The cognitive attributes are valued by implementing *Algorithm 1* for CTA and use-case task associations are defined by implementing *Algorithm 2* for HTA. Also subsequent changes in database proxy and CAIRIS DTDs are made.

5.2.3 Python Scripting and Graphviz Models

Finally, the Python scripting for generating graphical HTA models for use-cases and tasks is done. The model definition concepts from previous CAIRIS models and *Graphviz* are used for creating HTA models. Also, the test case that adds, updates, deletes and retrieves model objects is created. The purposes of this test case is to ensure database changes, stored procedures and methods.

5.3 Case Study - ERTMS Signaller

The aim of the case study is to validate process-techniques and tool-support for safe, secure and usable design framework. Using the analysis from this case study, the extension of design framework to incorporate the human factors engineering techniques is proposed. Also, this is tool-supported by implementation in CAIRIS.

For preliminary evaluation of use-case specifications informed TA approach using CTA and HTA is applied in rail infrastructure. The approach is going to help identify human factors issues in the form of levels of human failures. These issues will point out tasks which needs more attention from human factors experts. Using this approach the potential safety hazards and associated security risks will be identified. Hence these security risks, safety hazards and human factors issues within tasks will help to achieve safe, secure and usable design solutions.

5.3.1 Overview

Nowadays, due to information and technological advancements in critical infrastructures the jobs (task routines) are becoming more centred around mental (cognitive) abilities as compared to physical. As with the implementation of the ERTMS operational concept in rail, the working relationship is more dependent on team work and coordination capabilities. For example, the train operators work in conjunction with each other to ensure safe,

secure and efficient train operations. Thus, ERTMS specifications are used to conduct TA for the role of *Train Signaller*.

5.3.2 Task Breakdown

A rough profile of *A Day in the Life of a Train Signaller* is sketched, which consisted of task breakdown in a time-line from 0030 to 2350 hours as follows:

1. **0030:** Signaller combines signalling control areas (workstations) as planned, for granting *Possessions and Isolation*, to be worked by one signaller.
2. **0045:** Signaller grants *Possession and Isolations* using automatic blocking facility. Isolated areas are automatically blocked to electric traction via train describer functions, recorded automatically in operations log/ telegram journal.
3. **0530:** Signaller reverts as planned to default control areas for normal service delivery, to be worked by one signaller.
4. **0545:** Monitor trains running normally on ERTMS for *Stock Positioning* during morning peak service.
5. **0720:** Monitor trains running on ERTMS with some slight perturbations, and delays for *Conflict Prediction and Resolution* options. These options include short termination, forming stock short, and re-platforming etc. Signaller initiates *Regulator Intervention* which is also limited to selecting option/s, where some trains are manually routed.
6. **1045:** Signaller grants *Off-peak Blockage* of station platforms for cleaning. Using electronic lockouts which are recorded automatically in operations log/ telegram journal.
7. **1330:** Signalling faults known as failure points are indicated via alarm system. Conflict prediction and resolution functionality flag deviations from normal timetable until faults are rectified.
8. **1600:** *Broken Rail* on up-line requirement for traffic to be worked over one line until broken rail is temporarily plated. Conflict prediction and resolution functionality of ERTMS provides modified temporary timetable (involving cancellations) until faults are rectified. Signalling control area requires to be split in order to reduce contingent workload whilst one Signaller deals with emergency only scenario/s.
9. **2000:** Planned test of back up facilities is conducted using traffic simulators. This provides assurance of hot stand-by, and maintains competencies.

10. **2200:** Special freight train requires running two functions, which are *Route Availability* for safety and *Sectional Running Time* for performance. This is verified by *Operational Planning* functions of ERTMS.
11. **2350:** Signaller calls up list of standard possession and isolations, ahead of service run down, to establish real time *Order of Implementation*.

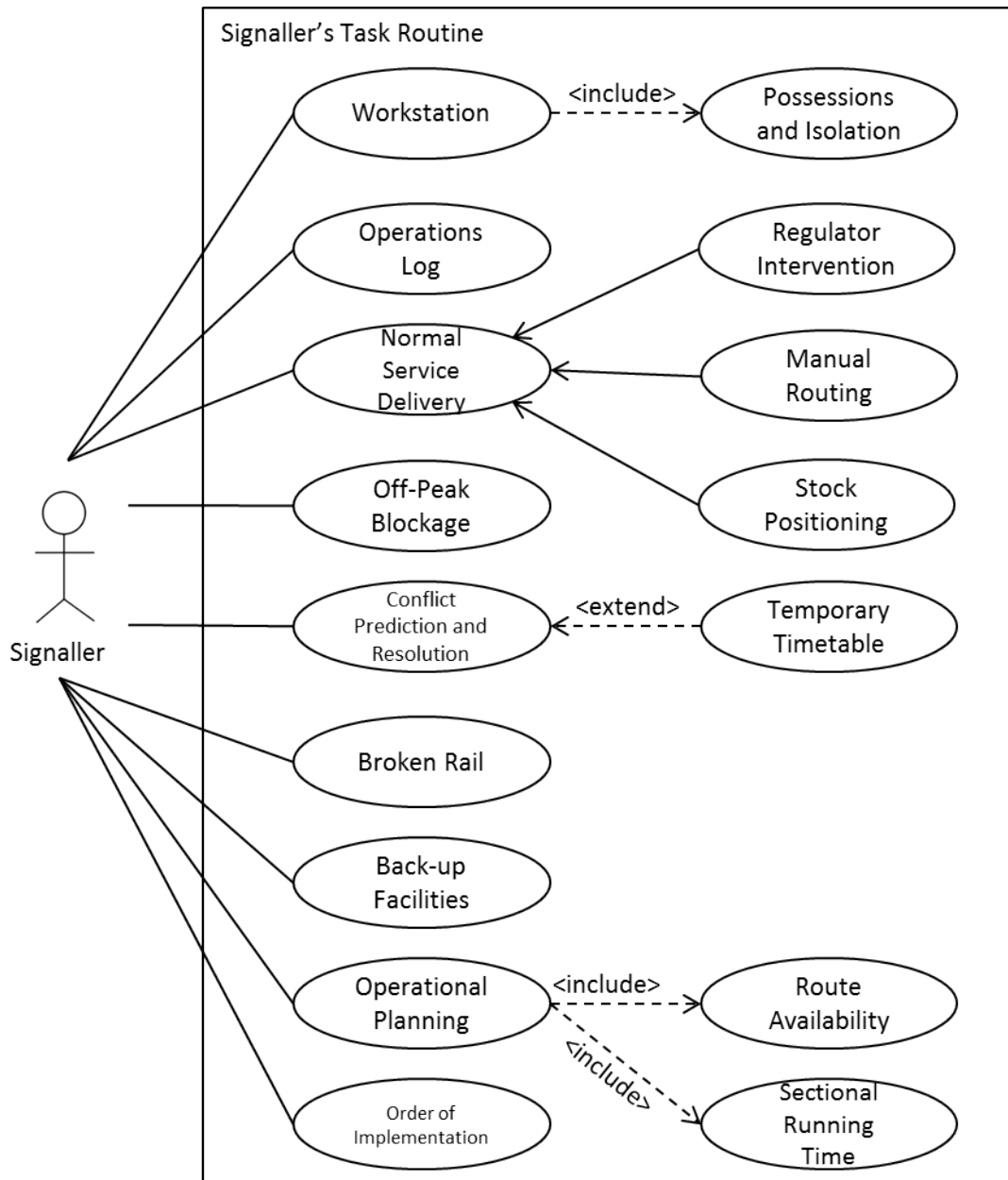


Figure 5.4: Use-Case View for ERTMS Signaller

5.3.3 Use-case View

For explanation purposes, 'ERTMS – Signallers Task Routine' is modelled where the main actor is defined as *signaller* which is interacting with different use-cases of system

to complete its daily job. The peak hours of operation are dependent on region and type of traffic. However, the peak hours can be generalised between 0730 to 0900 hours and 1630 to 1800 hours. This is shown as use-case view in Fig. 5.4.

The use-case view comprises of three main components:

1. **Actor:** An actor is external process, person or thing interacting with system. For example, train signaller.
2. **Subject:** A subject is defined as a classifier for use-case. For example, *workstation* which needs to be combined.
3. **Use-Cases and Relationships:** A use-case is a unit of system functionality. There are four kinds of use case relationships including association, extend, include and use-case generalisation. For example, the *Operational Planning* use-cases's relationship with two other use-cases.

5.4 Task Analysis

5.4.1 Personas for Task Elicitation

The ERTMS Operational Concept was used to develop an understanding of the job of Train Signaller. The open-source documentation and literature specified in Table 5.2 was used to ground knowledge. This knowledge was supplemented by interviewing a number of other relevant rail stakeholders. A total of 4 interviews were conducted, one from human factors expert with focus on TA methodologies, one from safety engineer for potential hazard analysis using human-error sources and two from train signallers for collecting data about ERTMS signalling tasks performed in routine.

Table 5.2: Documentation and Literature used for Train Signaller Personas

| Ser. | Article Title | Author | Publisher |
|-------------|--|---|---|
| 1. | A Day in the Life of a Train - Operational Concept | ERTMS | Operational Principles and Rules - Technical Document |
| 2. | Network Rail - Signalling Control Centers | Network Rail | Published and Issued by Network Rail - Module A5-5 |
| 3. | Operational Concept for The European Railway Traffic Management System | Rail Safety and Standards Board | RSSB-ERTMS-OC Issue 2 |
| 4. | Understanding Railway Signaller Tasks and Operations | Ex-Signalman and Human Factors Consultant | Interview Notes |

During this process models associated with the role of train *Signaller* were defined.

From knowledge base, 73 factoids were elicited, which grounded 11 argumentation models for the persona of a train signaller (*Daniel*). These argumentation models contributed towards the narrative of Daniel, explaining his activities, attitudes and aptitudes. Using personas narrative for *Daniel*, 16 major tasks were elicited for the role of train signaller. For example, the task of *Combine Workstation* is found from persona characteristic of activities for *Daniel* as shown by highlighted text.

*Daniel is performing the job of railway signaller. Daniel working from his signaller's workstation is responsible for monitoring and controlling train movements after **combining workstations**. He has to manage the signaling control operations for train movement, in order to ensure safety of people.*

These tasks were organised in rough tabular form and fed back to stakeholders for validation. The stakeholders responded back with comments valued as review and feedback.

5.4.2 Use-Case Specifications Informed Task Analysis

Use cases were identified and specified, using a pre-defined format. Using *Microsoft Excel*, points were scribbled down along-side data collection. This was an iterative process, where each use case specification went through series of transformations.

| | | |
|---|---|--|
| Use Case Title | Conflict Prediction and Resolution | |
| Abbreviated Title | Conflict and Resolution | |
| Use Case ID | UC_9 | |
| Actors | Signaller | |
| Objective: User desires to predict capacity of traffic management of the ERTMS, using conflict and resolution functionality. | | |
| Pre-Conditions: User points out failures indicated via alarm systems. For example, an over-crowded terminal station etc. | | |
| Task Sequence: 1. Use case starts when user wants to predict operation conflicts. 2. User monitors centralised traffic control system. 3. User detects potential operation conflicts. 4. User suggests optimal scheduling strategies for delays and deviations from timetables. 5. Use case ends. | | Exceptions: User fails to make timely predictions due to heavy work load and stress. |
| Post-Conditions: User provides advance plat-forming/ routing options to minimise delay using Automatic Route Setting (ARS). | | |

Figure 5.5: Use-Case Specification for 'Conflict Prediction and Resolution'

There were three major parts for each use-case: actor (performing the task), steps (task sequence) and conditions (identifying constraints/ exceptions). After careful consideration, a total of 16 use case specifications were specified. For example, Figure 5.5 specifies a use case for *Conflict Prediction and Resolution*. Following validation from stakeholders, these use case specifications were imported into CAIRIS.

5.4.3 Cognitive Task Analysis

After specifying the use cases, CTA was conducted by scoring each use case against cognitive reactions. For example, in the use-case of *Conflict Prediction and Resolution*, the values assigned were as follows: vigilance was *High*, situation awareness was *Medium*, workload was *High*, stress was *High* and risk awareness was *Medium*, with a defined rationale where *under manual control train movements or alterations in timetable may cause additional workload*. These values of cognitive reactions were fed into the *Algorithm 1*, where the mean was calculated as 3. This indicated that the *Conflict Prediction and Resolution* use case was associated with a *High* level of human failure. Consequently, the design analysis of this use case lead to situations where there is a strong tendency towards mistakes or errors.

Table 5.3: Cognitive Task Analysis for Use-Case Specifications

| <i>Use-Case ID</i> | <i>Use-Case Name</i> | <i>Vigilance</i> | <i>Situation Awareness</i> | <i>Workload</i> | <i>Stress</i> | <i>Risk Awareness</i> | <i>Level of Human Failure</i> |
|--------------------|---------------------------------|------------------|----------------------------|-----------------|---------------|-----------------------|-------------------------------|
| UC-1 | Combine Workstation | Low | High | Medium | Null | High | Medium |
| UC-2 | Grant Possessions and Isolation | Medium | Medium | Low | Null | Medium | Low |
| UC-3 | Maintain Operations Log | Low | Medium | Low | Low | Low | Low |
| UC-4 | Ensure Normal Service Delivery | Low | Medium | Medium | Low | Low | Low |
| UC-5 | Monitor Regulator Intervention | Low | Low | Medium | Medium | Low | Low |
| UC-6 | Conduct Manual Routing | High | Low | Medium | Low | High | Medium |
| UC-7 | Plan Stock Positioning | Low | Low | Medium | Low | Low | Low |
| UC-8 | Grant Off-Peak Blockage | High | Medium | Medium | High | High | High |
| UC-9 | Conflict Predict and Resolution | High | Medium | High | High | Medium | High |
| UC-10 | Issue Temporary Timetable | Medium | Medium | Medium | Low | High | Medium |
| UC-11 | Identify Broken Rail | High | Low | Low | Medium | Medium | Medium |
| UC-12 | Test Back-up Facilities | Medium | Medium | Low | Low | Low | Low |
| UC-13 | Map Operational Planning | High | Medium | Medium | Low | High | Medium |
| UC-14 | Run Route Availability | High | High | Medium | High | Low | Medium |
| UC-15 | Run Sectional Time | Low | Low | Low | Low | Low | Low |
| UC-16 | Order of Implementation | Medium | Medium | Low | Low | Low | Low |

The different values of cognitive reactions for the use cases is shown in Table 5.3, together with mean calculation from *Algorithm 1*.

5.4.4 Hierarchical Task Analysis

With the help of *Algorithm 2*, the colour coded HTA graph was generated with use cases of Low, Medium or High level of human failures, as shown in Figure 5.6. Here, in the

HTA graph, 9 use cases and tasks can be seen impacting each other. Based on the full HTA graph, 3 use cases – *Combine Workstations*, *Grant Off-Peak Blockage* and *Conflict Prediction and Resolution* – correspond with *High* levels of human failure.

By conducting TA as a combination of CTA and HTA tools, the cognitive load on humans parallel to hierarchy of tasks is better understood. For example, the use case *Map Operational Planning* depends on *Run Route Availability* and *Run Sectional Time*, where cognitive reactions like *vigilance*, *situation awareness* and *workload* are important. This breakdown highlights tasks dependency and logic behind goals, whereas resources, time and expertise are evaluated using cognitive reactions. Both equip human factors experts with sufficient knowledge when making design decisions.

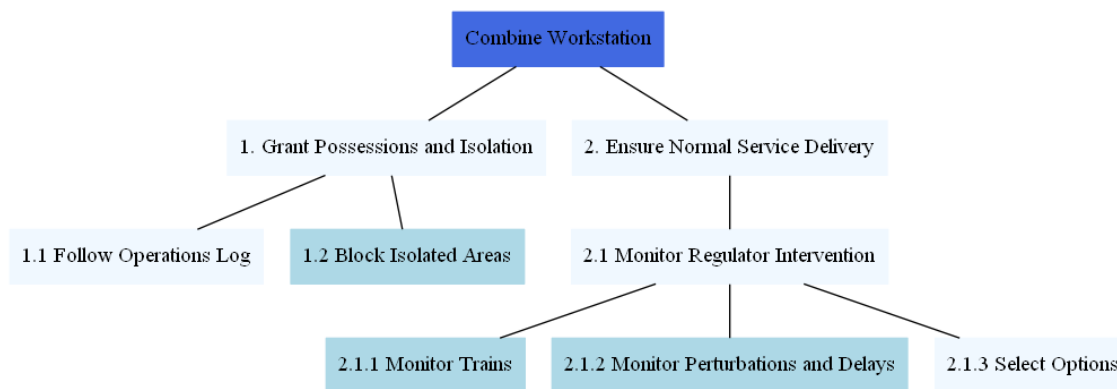


Figure 5.6: HTA Graph with Levels of Human Failure

5.4.5 Risk Analysis

During the specification of *Conflict Prediction and Resolution*, an exception was identified where a user fails to make timely predictions due to heavy workload and stress. This might occur due to the vulnerability of *Lack of Independent Check*, where the user should update checklists with timely prediction data. This vulnerability affords two threats: *Delays during Routing* and *Operational Conflicts*. These threats contribute to the risk *Failure of Automatic Route Settings*, and this failure leads to hazard *Collision between Trains*, with severe consequences.

5.5 Discussion

The approach entails TA as the human factors technique for determining potential human error sources. These human error sources highlight possible security risk elements in the form of vulnerability, threat, risk, and hazard. The intent and effort are recognised using CTA, where attributes like vigilance, situation awareness, workload, stress, and risk awareness are contributing factors. However, the task and use case hierarchical

breakdown using HTA contributes to an understanding of the division of effort between these tasks.

This is an area where human factors experts can provide important feedback. When presented with human error sources behind tasks performed, they can benefit from a graphical visualisation to show the tasks requiring more attention. By collaborating with security and safety engineers, the potential hazards arising from these tasks are also visualised using threat modelling and risk analysis in CAIRIS. Here, CAIRIS developed the link between tasks identified for TA and vulnerabilities resulting from these tasks.

With the occurrence of exceptions, possible exploitation opportunities are identified. For example, in this case study, the major exceptions found in the use cases are *power failure, equipment failure, conflicts and delays, track circuit failure, etc.* These exceptions link to KAOS goal models, which give security and safety experts an idea about possible vulnerabilities leading to threats, risks and hazards (i.e. risk analysis). Similarly, the cognitive reactions defined against each use case could determine the potential human error sources using HFACS framework.

Using HFACS, each use case with the highest level of human failure is labelled against the closest possible description of human error. For example, the use case *Conflict Prediction and Resolution* corresponds to a high level of human failure, where vigilance, workload and stress are important. Hence, the chances of occurrence of *Skill-based Error* and *Violation* are high, requiring scrutiny from human factors experts. Vigilance and workload may lead to the identification of *Decision Error*, but this is unlikely because this type of error results from a wrong judgement during emergency situations, rather than during routine operations.

5.6 Summary

In this chapter, an approach where use cases drive TA for designing and evaluating safe and secure rail infrastructures is presented. The rail infrastructure for design analysis is catalogued and, through a preliminary evaluation on regular tasks performed by an ERTMS Signaller, human error sources behind these tasks are highlighted. In doing so, it is shown how these human error sources contribute towards design solutions by identifying safety hazards and security risks.

In presenting this approach, three contributions have been made. First, a TA approach is derived from the security and requirements engineering IRIS framework using concepts such as roles and personas, task and goal-obstacle modelling. Second, it has shown how CTA and HTA can be combined as single, tool-support TA approach to highlight the importance of mental load with a detailed task breakdown. Finally, it has shown how use case specifications assist with task sequencing and exception identification. These exceptions help security and safety experts to conduct risk and hazard analysis by

identifying potential vulnerabilities and threats hidden beneath system design.

TA with CAIRIS as tool-support facilitates other kinds of analysis, including asset, goal-obstacle, responsibility, threat, and risk modelling, and even hazard investigation using safety analysis techniques. Thus, by using this approach the human factors experts are given a chance to work in collaboration with security and safety experts to analyse and make collective design decisions for critical infrastructures. In the next case study, this approach is built by integrating safety analysis techniques and methods to further facilitate the design of safe, secure, and usable rail solutions.

Chapter 6

Integrated Design Framework for Facilitating STPA

In this chapter, the research background stating the processes and techniques recognition for facilitating STPA model using design framework is presented. The design framework is applied in the final case study of 'Cambrian Railway Incident'. This is followed by incident overview with breakdown of events. Finally, the application comprising of identification of role and personas, task model, and use-case specification for safety analysis using STPA is conducted. This also initiates the basis for risk analysis which helps to draw recommendations for the incident. Hence, stating that the proposed design framework is an exemplar for resolving safe, secure and usable design solutions in rail.

6.1 Safety Analysis

STPA is used to identify control actions and causal factors behind accidents to improve system design (Leveson 2018). The approach revolves around a series of pre-defined steps followed by experts. Using STPA analysis, the security controls can help to mitigate security risks and potential safety hazards. For example, poor design decisions may lead operators to make human errors or mistakes where rules are un-intentionally disobeyed (Lahoz 2015). Using STPA the relevant human cognitive processes are modelled, as means of fleshing out unwarranted assumptions. Consequently, the system safety and security may be compromised due to human intervention in the form of errors or violations.

IRIS framework has been used to identify security risks leading to safety hazards for identifying human factors issues (Altaf et al. 2019). This is achieved by identifying and modelling assets associations, roles and personas, vulnerabilities, threats, risks, tasks and goals (Faily 2018). Based on the IRIS framework and complementary 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 (Altaf et al. 2021). These failure levels are used to identify associated safety and security design solutions by identifying potential hazards.

An extended design framework has been 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.

6.2 STPA Process Model

In critical-infrastructures, a higher percentage of design decisions are centered around safety, whereas the consideration for human factors is less frequent. System safety is usually compromised due to human intervention in the form of errors and mistakes (Reason 1990). The safety analysis of system is responsible for timely identification of all such potential hazards which arise from security risks and human failures or errors, but also lead to accidents.

The STPA process model 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.

6.2.1 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 (Atzeni et al. 2011). 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 un-

derstand 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 (Altaf et al. 2021).

6.2.2 Step 1: Accident, Hazard and Constraint

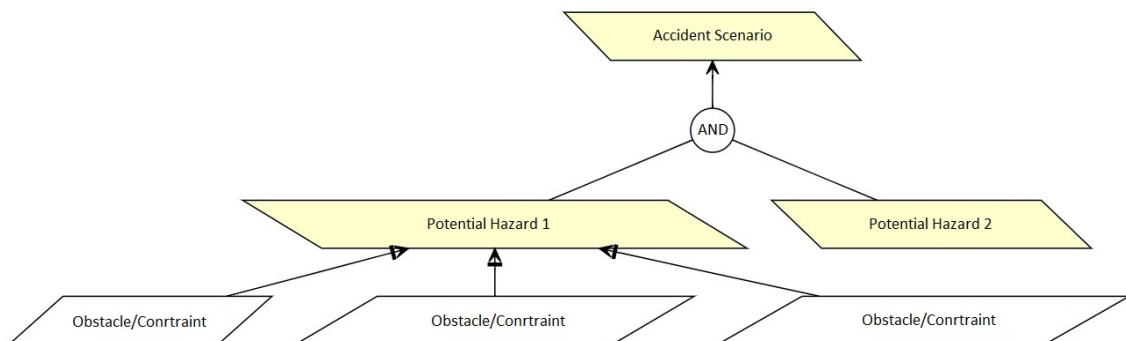


Figure 6.1: Accident, Hazard and Constraint Model using KAOS Association in CAIRIS

The STPA process begins by defining the accidents (losses) in relation to identified hazards (Leveson 2018). 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 as shown in Fig. 6.1, provides more meaning and understanding for further analysis by domain experts.

6.2.3 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 (Leveson 2018). An effective way for modelling these control structures within CAIRIS is by using DFD as shown in Fig. 6.2. 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.

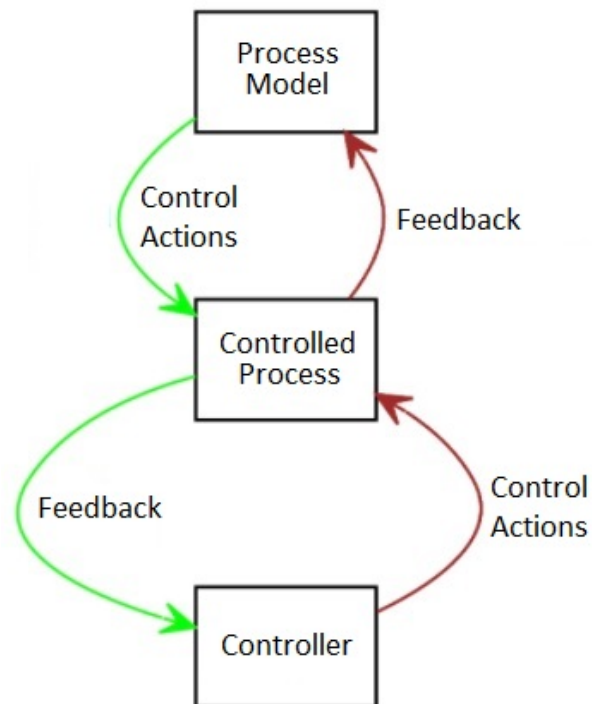


Figure 6.2: Model Control Structure using DFD in CAIRIS

6.2.4 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 (Leveson 2018). In CAIRIS, an unsafe control action is presented using *obstacle* and the safety constraint is modelled by associating these obstacles with DFDs.

6.2.5 Step 4: Causal Factor

The causal factors are identified by analysing the controllers, processes, feedback and control paths (Leveson 2018). In CAIRIS, the identified tasks during human factors analy-

sis, 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.

6.2.6 Step 5: Risk Analysis Model

These identified causal factors are also defined as system vulnerabilities leading to hazards (accidents). The vulnerabilities are 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.

6.3 Case Study - Cambrian Railway Incident

The aim of the case study is to present design framework as an exemplar for resolving safe, secure and usable design issues in rail. The process behind STPA is facilitated by using security and human factors engineering approaches. This motivation aims to highlight process-requirements for STPA by tool-supporting it using CAIRIS.

The real life incident of *Cambrian Railway* is used to conduct a case study based on qualitative evaluation of proposed design framework. The incident took place on 20-Oct-2017 on the Cambrian Coast Line, where a train faced over-speeding because of technological failure (Les 2019). The train was following the route of Cambrian Coast Line as shown in Fig. 6.3.

6.3.1 Overview

During service between Barmouth and Llanaber, the train was travelling with three times its actual speed. The over-speeding was timely observed by its train driver and he immediately reported the fault to concerned authority. After that manual routing was conducted by the train driver and signaller, until the rectification of fault. 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 (RAIB 2019).

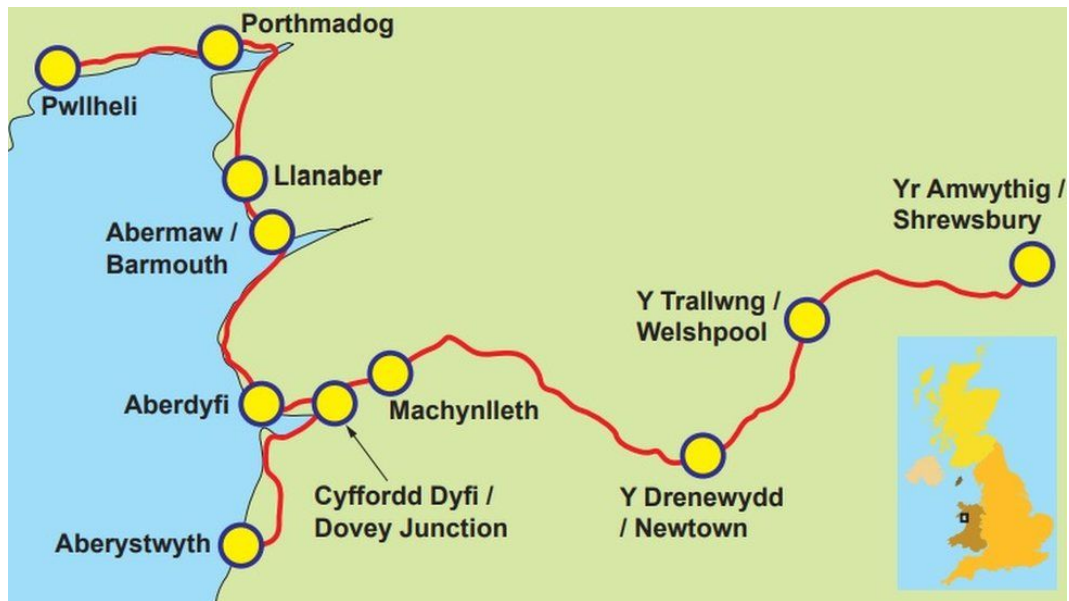


Figure 6.3: Route for Cambrian Coast Line (Les 2019)

6.3.2 Breakdown of Events

The following breakdown of events is used to better understand the scenario.

Automatic Computer Restart On the evening of 19-Oct-2017, an automatic computer restart was scheduled. The main purpose of this restart was to upload Temporary Speed Restriction (TSR) data on signalling computer.

Independent Check by Signaller As per routine, the signaller was supposed to ensure the data about TSR is correctly uploaded on signalling computer. The display screen showed an incorrect upload of data, but no independent check was performed by signaller before service.

Passenger Train Service On the morning of 20-Oct-2017, the train driver prepared train for service after roll-over. The first passenger train service was initiated at 0717 hours.

2J035 Train Service The issue remain unidentified by three passenger train services. At 0852 hours, the fourth passenger train service 2J035 was on its route from Machynlleth to Pwllheli.

Over-speeding of 2J035 At around 1002 hours, the train was on its route from Barmouth to Llanaber. The train driver observed that the train is travelling at three time its actual speed i.e., 80 km/h (50 mph) rather than 30 km/h (19 mph).

Fault Reporting After observing this over-speeding and incorrect data on Driver Machine Interface (DMI). The train driver reported this fault to signalling technician.

Manual Routing The train driver and signaller reverted to written and verbal communication to ensure service without disruption.

6.3.3 Choice of Incident

The choice of this incident is based on multiple factors like signalling system, service type, form of rail transit, and design implementation. The Cambrian Coast Line was following the *ERTMS*. ERTMS is based on ETCS as a rail signalling system, which ensures reliability, optimised capability and automation. The achievement of these qualities in ERTMS is dependent on safe, secure and usable design goals. The service type is *Passenger Train* which is life 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. By design implementation, it meant *Information Technology* application, in order to keep up with latest design requirements and trends.

6.4 Partial-STPA Assessment

The Cambrian Incident case study application of proposed integrated design framework begun by collection of data. All open source documentation and literature was collected and surveyed as shown in Table 6.1.

Table 6.1: Literature Survey on Cambrian Incident

| <i>Ser.</i> | <i>Article Title</i> | <i>Author</i> | <i>Publisher</i> |
|-------------|--|--|--|
| 1. | A Day in the Life of a Train | European Railway Traffic Management System | Operational Principles and Rules Version 5 - OPS.117 |
| 2. | Cambrian ERTMS Loss of Temporary Speed Restrictions | Ian Mitchell | IRSE Publication - March 2020 |
| 3. | Lessons Learnt over Train Speeding on Cambrian Line | British Broadcasting Corporation | BBC News Article |
| 4. | Loss of Safety Critical Signalling Data on the Cambrian Coast Line | Rail Accident Incident Branch | Incident Investigation Report |
| 5. | Signalling Control Center - Technical Document | Network Rail | Published and Issued by Network Rail Module A5-5 |

The basic information such as project name, background data and goals of case study were initiated within CAIRIS, in order to formally start data collection and organisation. Moreover, the relevant stakeholders were determined including safety, security and human factors experts. The stakeholders were categorised into *scriber, facilitator and participant*. In CAIRIS, the environments were defined using name and description.

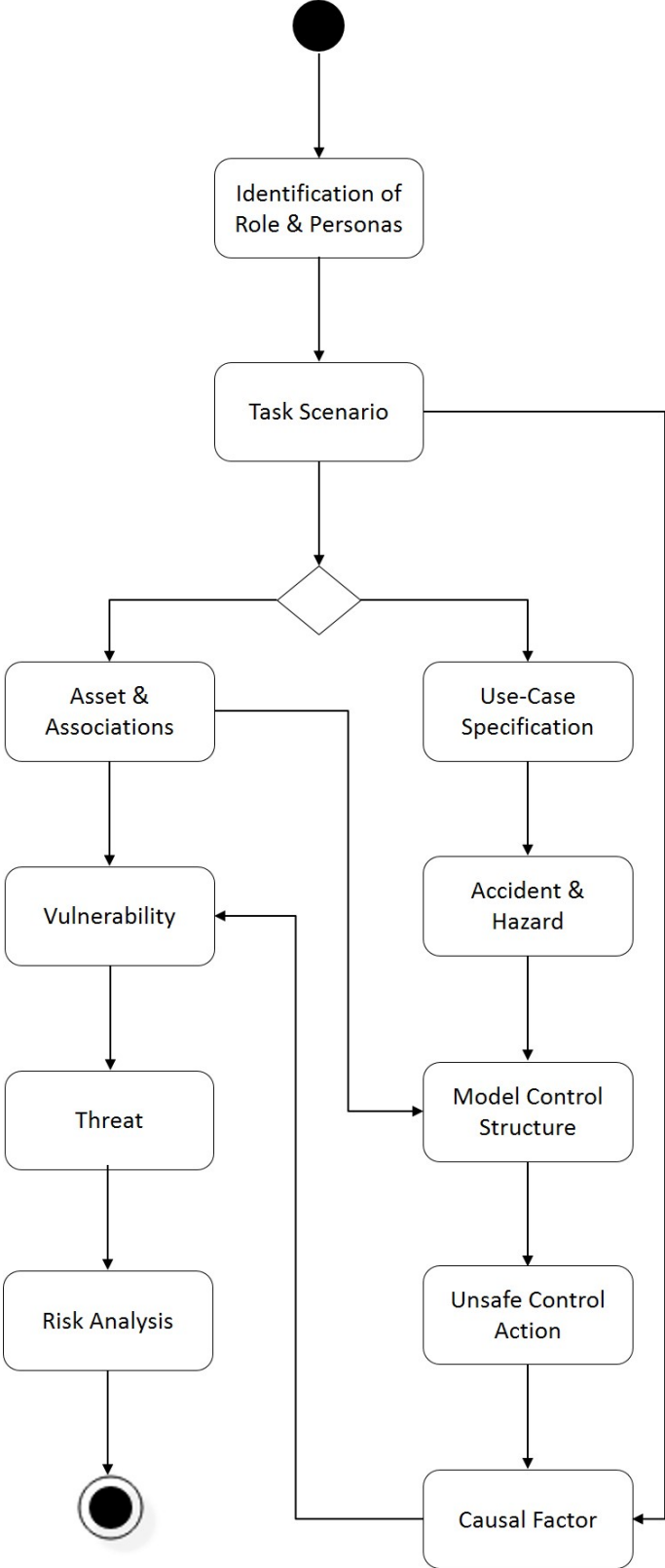


Figure 6.4: Activity Diagram for STPA Using Design Framework

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 application was completed in three stages. This was shown using activity diagram in Fig. 6.4. During stage 1, the human factors approaches such as identification of role and personas, task modelling and use-case specifications were conducted. The safety analysis using STPA was done in stage 2 where accident, hazard and constraints were defined, control structures were modelled, and unsafe control actions were determined including causal factors. Eventually, leading to stage 3 where security techniques based on attacker, vulnerabilities, threats and risk analysis were managed.

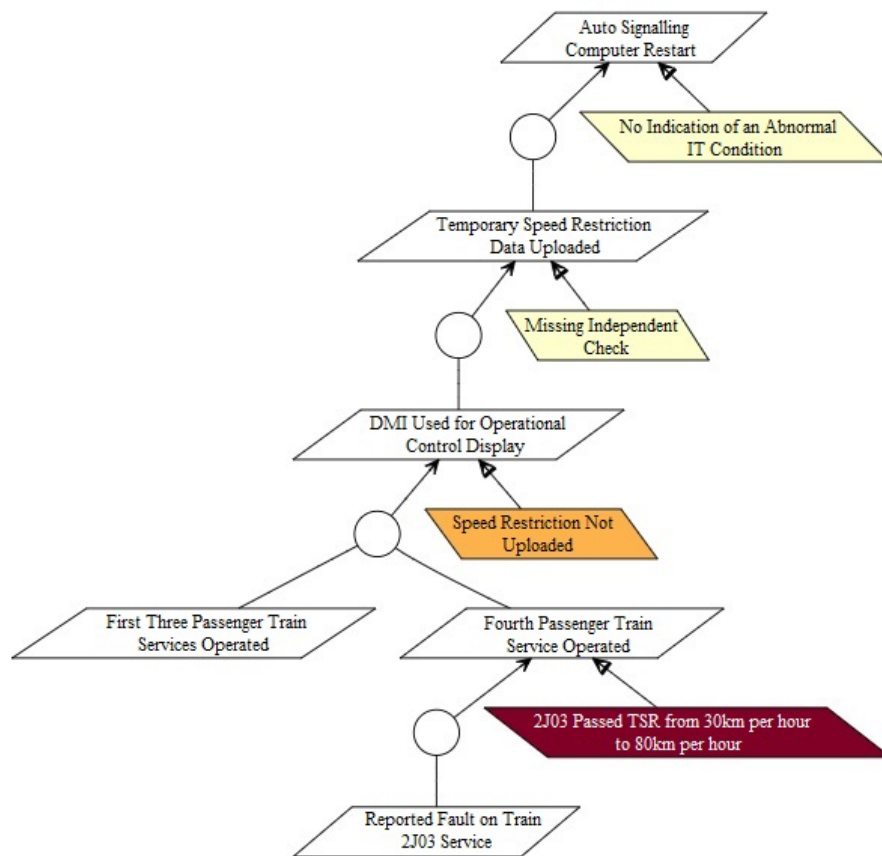


Figure 6.5: Goal-Obstacle Model for Cambrian Incident Case Study

The Cambrian Incident case study was modelled using KAOS to show a general scenario behind the accident (RAIB 2019). For this purpose, 6 goals and 4 obstacles were identified and their associations were defined as shown in Fig. 6.5, 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 *TSR Data Uploaded*, where the obstruction was caused due to *Missing Independent Check*. The TSR data was displayed on DMI avail-

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

6.4.1 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 performing 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.

Summary Activities Attitudes Aptitudes Motivations Skills Contextual Trust Intrinsic Trust

As a part of his routine, Neil is supervising train drivers. The train drivers precede their journey, based on directives known as movement authority from signallers. Neil plans and sometime reschedules the journey information based on ERTMS standard which becomes visible to train drivers on their on-board DMI.

Neil also keeps record of GSM-R status of every train. In case, of any incident the investigation can be conducted using the traffic information kept as evidence by Neil.

+ Environment

Peak Hours Off-Peak Hours

Roles Narrative

Direct User

| | Role |
|---|-----------|
| + | |
| - | Signalman |

Figure 6.6: Persona Characteristic of 'Attitudes' for Neil in CAIRIS

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. For example, consider the following narrative of persona characteristic of *activities* for *Ray*:

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 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.

Ray is managing his operations by the help of the DMI available in his driving cab. The DMI is based on ERTMS. Through DMI, Ray enters data into an on-board system. As a result, Ray receives information about his train route and allowed train speed. In order to ensure safe movement of trains, Ray takes action based on movement authority sent to him by signaller.

There are some activities which Ray performs to ensure efficient movement of trains, like switching of isolation mode. It is Ray's job to identify the End of Authority for trains. Ray's driving cab has ERTMS reset, and sometimes under special circumstances he has the authority to reset. It is also Ray's job to keep on observing signals for danger and stop the train immediately to avoid any catastrophic accidents.

Figure 6.7: Task Participation Form for 'Self-Test Function' in CAIRIS

Similarly, the persona of *Neil* was based on 18 argumentation models. These argumentation models were used to understand persona characteristics which form the narrative for personas as shown in Fig. 6.6. This narrative backed-up by factoids from document references lead towards identification of task models for further analysis.

A total of 19 tasks were created in CAIRIS, where 11 were derived from the persona of *Ray* and 8 from *Neil*. Tasks were defined using narrative, participants, dependencies, consequences, benefits and concerns within an environment. The narrative helped to understand the objective of task along with description of procedure. However, the persona play a major role behind the recognition of tasks. A task participation form was used to relate persona with task using usability attributes like duration, frequency, demands and goal conflict.

For example, the task of *Perform ETCS Self-Test Function* was found from persona characteristic of activities for *Ray* as shown by highlighted text.

*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 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.***

Using *Task Participation Form*, the required usability attributes were declared as shown in Fig. 6.7. The usability attributes with different values highlighted tasks with different colours during task models. These task models comprised of tasks against specified roles and personas. These models facilitated in specifying use-case actors and use-cases for human factors analysis. Also, these models helped to relate associated assets, threats and vulnerabilities, which assist experts during security analysis.

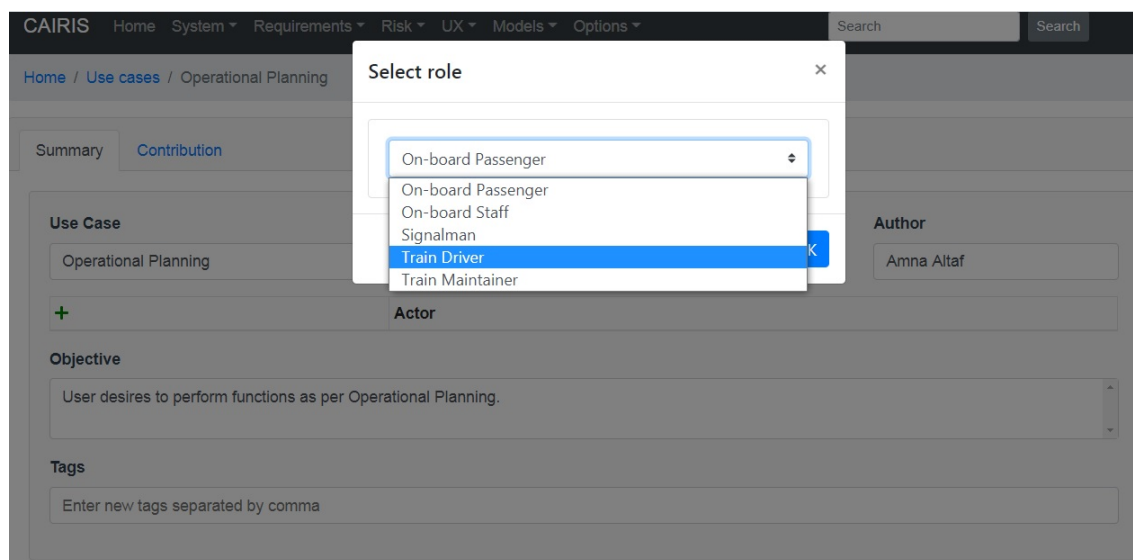


Figure 6.8: Actor Identification for 'Operational Planning' Use-case in CAIRIS

With the help of personas narrative and task analysis models, 17 use-case specifications were defined. Each use-case specification came with an objective, actor, pre-conditions, steps (task sequence), post-conditions and exceptions. For example, consider the use-case of *Operational Planning* as shown in Fig. 6.8. During task models, the use-case actors can also be linked with them showing relationship between role, persona, task and use-case. These elaborate models help experts to better visualise design of system along with specified environments.

6.4.2 Step 1: Accident, Hazard and Constraint

During TA and modelling, 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.

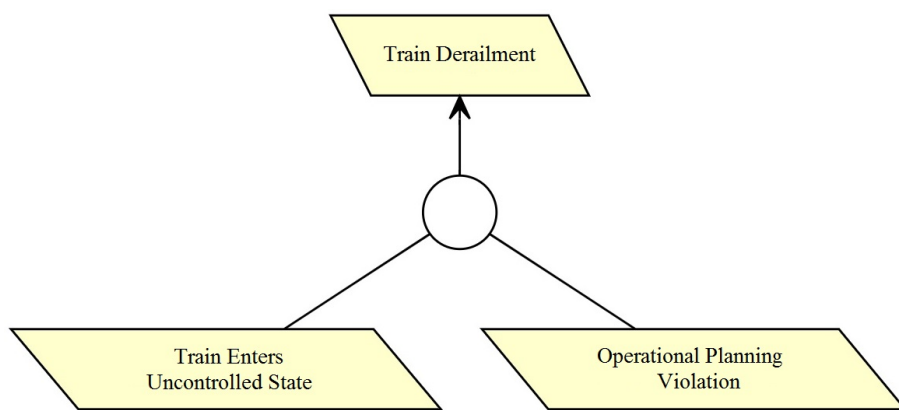


Figure 6.9: KAOS Association Between Accident and Hazard

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* as shown in Fig. 6.9.

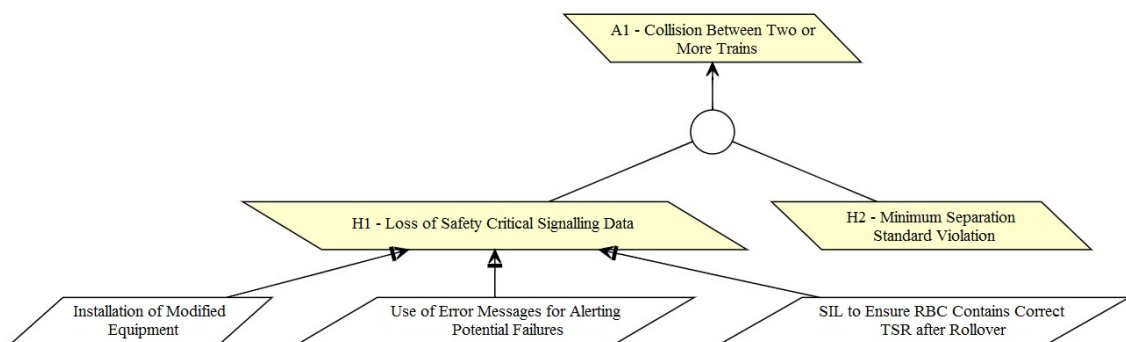


Figure 6.10: KAOS Association Between Accident, Hazard and Constraint

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. 6.10.

6.4.3 Step 2: Model Control Structure

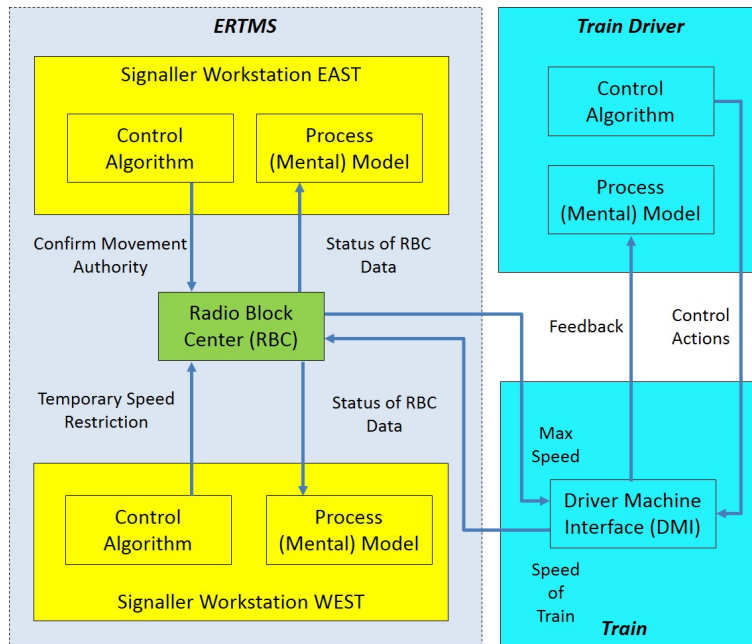


Figure 6.11: High-level Control Structure Model

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. For example, behind the DFD element of *Train Driver* there are control actions and feedback of information flowing between control algorithms of *DMI* and *Status of RBC Data*. The DFD in CAIRIS, shown in Fig. 6.12, was also used to construct high-level control structure model as shown in Fig. 6.11.

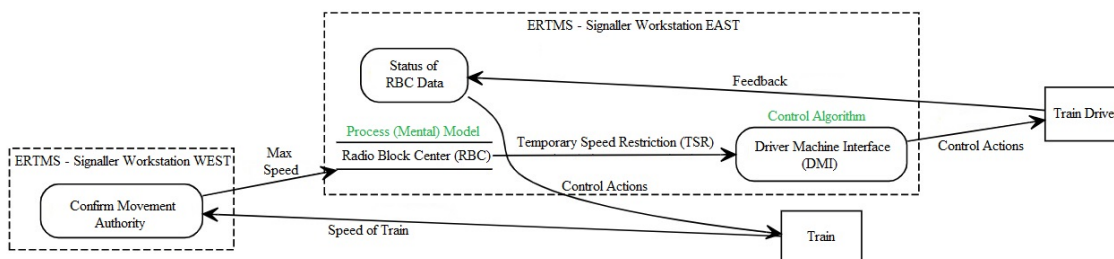


Figure 6.12: DFD of Control Structure Model using CAIRIS

6.4.4 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 6.2: 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 | |

6.4.5 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.

6.4.6 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 SIL* 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.

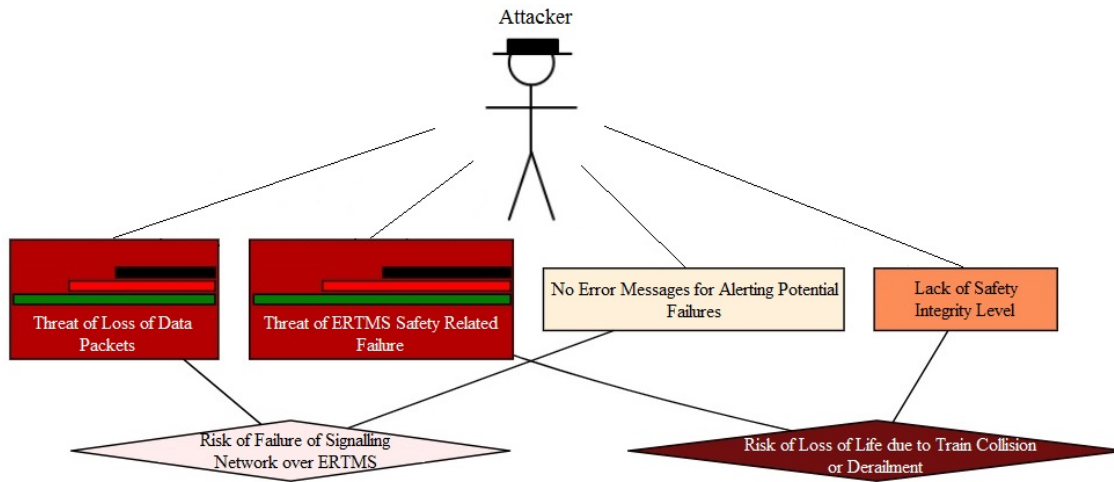


Figure 6.13: Risk Analysis Model Based on Attacker, Threat and Vulnerability

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. 6.13. 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.

6.5 Discussion

In this case study of Cambrian Railway Incident, STPA is applied in CAIRIS using human factors and security approaches. The human factors approaches such as identification of roles and personas, task modelling and use-cases are used to understand processes, asset associations and goal-obstacle models. In return, goal-obstacle models and DFDs (processes and datastores) are used to conduct STPA, where risk analysis based on recognition of attacker/s, threats, vulnerabilities, risks and misuse cases are done, simultaneously. All these process-techniques are tool-supported by open-source CAIRIS platform.

During auto signalling computer restart, the goal was to update correct TSR on DMI. But due to an IT failure, incorrect TSRs were uploaded. An extra check by signaller might have ensured correct data upload. Because the system lacked a task, where an extra check was to be performed. Using goal-obstacle model, an obstacle was raised where no notification was present during incorrect data upload or missing independent check. Hence, there was a requirement for additional safety assurance procedure in system

design, where missing independent check or incorrect data on DMI should have been timely noticed.

Once an issue was raised where the train was travelling three times its allowed speed. There was a requirement for accurate and effective safety management system, where enough data was available for tracing cause behind issues. Even these constraints were recognised as *Good Safety Management Engineering* and *Implementation of Defensive Programming*, against hazard of train entering an uncontrolled state. This potential hazard led to an accident where train derailment might have occurred.

There was a requirement for initiation of alert messages in case of issues. When implemented in system design should have notified incorrect or missing critical data on ERTMS signalling system. Eventually, raising the need for more robust configuration management where system integrity have been maintained by inclusion of warnings and alerts. This design requirement helped to minimise chances of potential hazard where there was *Operational Planning Violation*. During security analysis, risk of failure of signalling network over ERTMS occurred by exploiting vulnerability where no error messages were generated for altering potential failures.

An independent safety check was missing where the system design needed an extra SIL to ensure correct upload of data. This led to recognition of threat of loss of data packets. Therefore, essential system installation and configuration changes are recommended.

These all issues helped to understand an integration of concepts between safety and security, security and human factors, and human factors and safety. Thus laying a foundation of an overlap of concepts between IRIS framework, TA, HFACS and STPA, which leads to recognition of safe, secure and usable design framework. Using this integrated design framework, safety goals, security risks and human factors concerns were highlighted. Also, by tool-support the effort required by safety, security and human factors experts was minimised by providing automated and efficient design solutions.

6.6 Meta-Model of Design Framework

The process-techniques and tool-support from security by design approaches, human factors engineering techniques are used to facilitate safety analysis. Based on these methods, the IRIS framework, use-case specifications based TA, HFACS and STPA are presented as a meta-model and aims to act as an exemplar for resolving safety, security and human factors design issues in rail infrastructure as shown in Fig. 6.14.

This meta-model is a representation of linkage between three domain-specific approaches. The stakeholders (users) are also divided into three categories: security, human factors and safety experts. Each contributing towards their specific approach and as a result are able to visualise integrated models for better design and analysis.

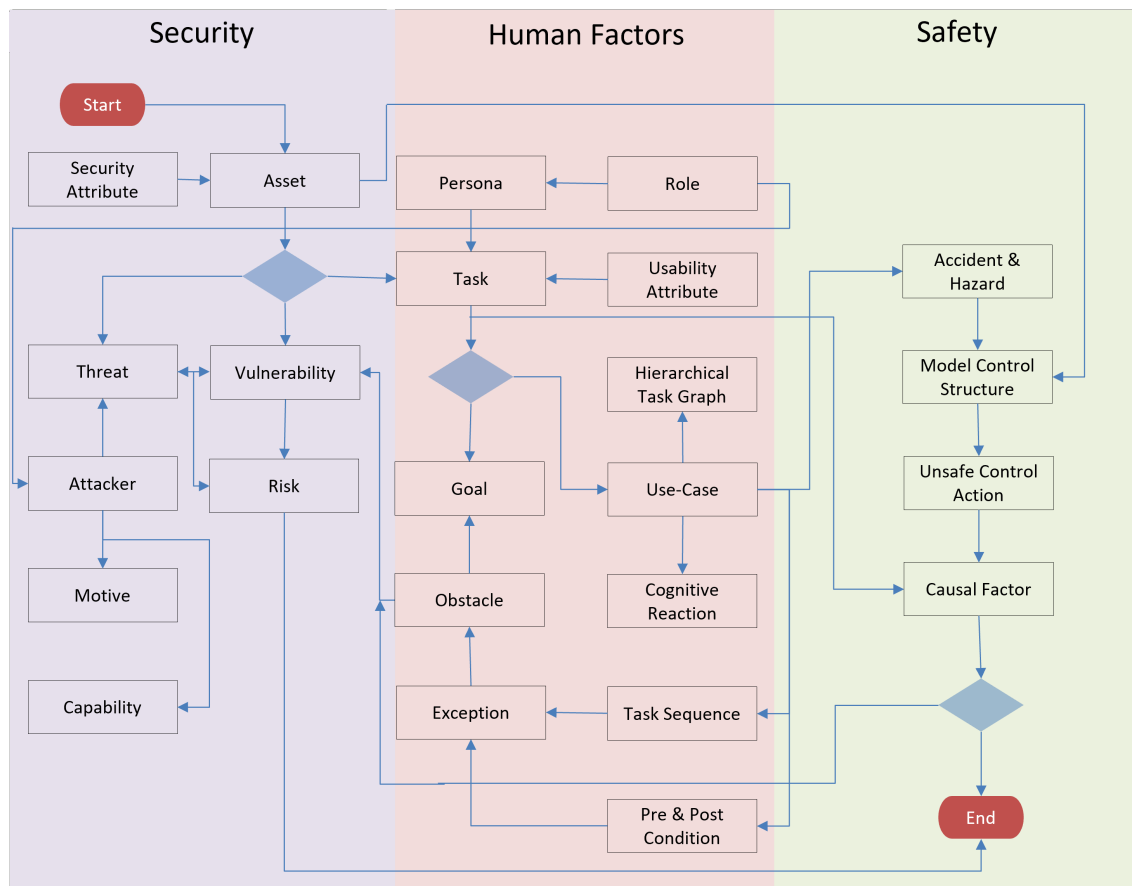


Figure 6.14: Meta-Model of Design Framework

The meta-model enables security experts to begin by identifying assets and their associations using IRIS framework and CAIRIS tool-support. At this stage, the security attributes namely, confidentiality, integrity and availability are valued. This leads to multiple paths. One way is to identify vulnerabilities leading to threats and risk analysis. Another is vice-versa where identification of threats contribute towards vulnerabilities and risks. Here, the perspective attacker with the motivation and capability are valuable. The human factors experts are responsible for identifying roles (attackers as well) and personas. This information helps the security experts to identify associated threats. Thus, the final option where the assets lead to task scenarios (human factors).

The human factors experts with the knowledge of assets and their associations, roles and personas describe tasks with usability attributes. This leads to the option of goal-obstacle modelling or a detailed TA. The goals are identified with potential obstacles leading to vulnerabilities (security experts) as well. As for TA is concerned, the use-case specification templates are filled up for performing CTA and HTA for evaluating human performance (error sources) using HFACS. The use-case specification comprise of task sequence containing a step-by-step actions with pre and post conditions. Here, the human factors experts might identify an exception towards these use-cases. These excep-

tions have the potential to contribute towards obstacles and vulnerabilities (risk analysis).

The use-case specifications by human factors experts lead to identification of accident and hazard. Hence, the safety experts are involved for STPA analysis. The STPA analysis begins by potential accidents and hazards enabling to model control structures. The assets listed by security experts also contribute towards modelling control structures. The worst case scenarios leading to hazards are recognised by defining unsafe control actions. The safety constraints are determined for minimising the unsafe control actions. Finally, the causal factors are identified by analysing the controllers, processes, feedback, and control paths. Using tasks from human factors experts, are linked-up with hazards and system-level constraints using KAOS goal refinement associations. Here, the security experts are facilitated as causal factors contribute towards identification of vulnerabilities, threats and risk analysis. Therefore, as a result of this meta-model all experts are given a chance to work together.

6.7 Summary

In this chapter, STPA process model is derived using the IRIS framework and CAIRIS platform. As a result, three signification contributions are made. First, this work has demonstrated 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 comprised of accident (loss), hazard, system constraint, component (control algorithm), process (mental) model, unsafe control action (obstacle) leading to causal factors. Second, this has shown 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 has 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 (Section 2.4.2). 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 have established grounds for understanding potential hazard scenarios and model better risk analysis. Hence, this research has built 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 contributed 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 helped to evaluate an integration of concepts between safety and security, security and human factors, and human factors and safety.

Chapter 7

Conclusion

In this chapter, the findings are used to analyse evaluation of research questions, as mentioned in Chapter 1. The results of this PhD thesis are concluded by summarising the key research findings and contributions. One important aspect is to determine how contributions from this research project are able to answer research questions, and resolve research challenges and limitations. Finally, future work directions are mentioned by presenting industrial perspective and viewpoint.

7.1 Evaluation of Research Questions

The aim of the PhD research is to identify a *design framework* for resolving safety, security and human factors issues in rail infrastructure. As a result of research gap identified in Chapter 2, three research questions were raised. These research questions are answered and evaluated as follows:

7.1.1 RQ1 - Integration of Concepts

The RQ1 raised in Section 1.3 was as following:

How the concepts from safety, security and human factors engineering can be integrated together to build the foundation for design framework?

First, the RQ1 is answered by the strong evidence provided by literature review in Chapter 2, where the alignment of concepts in terms of scope of integration, process-techniques and tool-support is available. The basics of safety-critical system and their design factors are used to develop a link with security-by-design techniques such as threat modelling, performance evaluation using human error recognition, and design techniques including personas. This is used to understand human factors engineering in terms of use-case scenarios, KAOS goal modelling language, and IRIS framework. Here, user-centered design approach of human factors using TA and their associated tools are used to overlap

with safety using HFACS framework for identifying human error sources leading to potential safety hazards. These safety hazards are analysed using STPA. This is also evident by the meta-model of design framework (Section 6.6).

Second, the case study application of 'Polish Tram Incident' has validated the integration of security with safety and human factors engineering. Here, three contributions are made in support of the raised claim. The asset modelling and their associations proposed using process-technique are identified by rail stakeholders where security attributes namely, confidentiality, integrity, availability are assigned. By using attacker personas the vulnerabilities are identified based on attacker perspective. This leads to the identification of threats with the support of scenarios and rationalises risks. These risks contribute towards safety hazards. On the basis of these potential hazards, the human error sources (active failures) are determined using HFACS framework. This aims to address and evaluate RQ1.

Finally, the security-by-design approach based on IRIS and HFACS framework with tool-support of CAIRIS have been successfully published in peer-reviewed Springer workshop. After blind peer-review the paper was published and presented to wide audience at workshop. This further strengthened the concept of integration of security with safety and human factors engineering. Also the rail stakeholders from Ricardo Rail gave feedback of the data and process, as they were closely involved during asset, role, task, goal-obstacle, requirement, dependency and risk analysis between humans and systems. Their human factors and safety expert contributed towards the design framework and case study selection as well. Generally, this PhD research has been based on *Empirical Evaluation* by case study application. The company has been involved during input and analysis of results based on feedback (email correspondence). Also, refer to Section 7.3 for more general reflection on validity concerns and stakeholders involvement.

7.1.2 RQ2 - Process-Techniques & Tool-Support

The RQ2 raised in Section 1.3 was as following:

How the processes and techniques based on safety, security and human factors engineering integration can be adopted, along with available tool-support options to propose a new design framework?

During Case Study I, the security-by-design approach was linked up with human factors techniques in terms of TA as a combination of CTA and HTA. By using CTA, the identified cognitive attributes responsible for affecting the task performance can further help to determine the human error sources using HFACS framework. Here, STPA was identified to be linked with IRIS framework to classify the identified hazards in more detail and determine the possible control actions for them. The identified hazards may be as a result of human errors or mistakes, for which HFACS framework was used. Therefore, establishing

a more refined process-technique and tool-support in terms of use-case specifications informed TA for security risk and safety hazard analysis. The TA as a combination of CTA and HTA has the tendency to identify tasks with mental workload (cognitive attributes) and structural breakdown of scenario.

The case study application of an 'ERTMS - Role of Signaller' helps to validate this formulated design framework as a combination of process-techniques and tool-support from safety, security and human factors engineering. A preliminary evaluation is done of regular tasks performed by an ERTMS Signaller, which highlights human error sources behind these tasks. This aims to address and evaluate RQ2.

Finally, the human factors engineering technique of TA based on use-case specifications template as a combination of CTA and HTA for identifying security and safety issues has been accepted for publication at a renowned conference venue. The work has been blindly peer-reviewed and based on feedback submitted for publication and presentation. Thus, helping to convey the methodology and findings to a wide audience which has further evaluated the RQ2.

7.1.3 RQ3 - Design Framework

The RQ3 raised in Section 1.3 was as following:

How by the application of proposed design framework the safety, security and human factors engineering design concerns are resolved in rail?

Using Case Study I and II, the concern of evaluation of design framework in rail infrastructure for resolving safety, security and human factors design issues is raised. This is implemented using Case Study III of 'Cambrian Incident' in which the design framework is presented as an exemplar for resolving all these design issues in rail.

In case study of Cambrian Incident, STPA method is applied using human factors and security approach. The human factors approach such as identification of roles and personas, task modelling and use-cases are used to understand processes, asset associations and goal-obstacle models. In return, goal-obstacle models and DFD (processes and datastores) are used to conduct STPA, where risk analysis based on recognition of attacker/s, threats, vulnerabilities, risks and misuse cases are done, simultaneously. All these process-techniques are tool-supported by open-source CAIRIS platform. Finally, the analysis from security-by-design approach (Case Study I) and human factors engineering techniques (Case Study II) contribute towards the identification of potential safety hazards (Case Study III). This security and human factors driven STPA is conducted using IRIS framework, use-case specifications informed TA, HFACS framework and CAIRIS as tool-support, leading to integrated *design framework*.

This helps to understand an integration of concepts between safety and security, security and human factors, and human factors and safety. Thus, laying a foundation of an

overlap of concepts between three domains, which leads to recognition of safe, secure and usable design framework. Using this integrated design framework, safety goals, security risks and human factors concerns are highlighted. Also, by tool-support the effort required by safety, security and human factors experts is minimised by providing automated and efficient design solutions. This case study aims to address and evaluate RQ3.

For evaluation of data and process behind design framework application, the rail stakeholders from safety, security, and human factors engineering including consultants from Ricardo Rail are presented with this study. A review of the STPA case study by a Senior Consultant summarised in a comment is quoted as:

"I like the use of the case study from a known and investigated incident, as it highlights where the CAIRIS tool can help in understanding that (or similar incidents)."

Also, the research paper written as a result of this case study has been submitted for publication at a renowned conference venue.

7.2 Key Research Findings

Based on research motivation in Chapter 1 and research gaps as pointed out by literature survey in Chapter 2, where there is a need for integration of concepts between safety, security and human factors engineering. This highlighted the requirement for mutual process-techniques and tool-support for formulation of design framework. Here the focus is on alignment of basic concepts such as asset, vulnerability, threat, risk and hazard for identification of design framework. Furthermore, there is visible lack of guidance to inform integration of different concepts, models, processes and techniques. Another research gap indicated the need to formulate task and goal-obstacle modelling techniques for achieving better safety and security, thus helping to capture the context required for safe, secure and usable design framework.

Therefore, this motivates rail industry wide need for identifying the alignment of concepts and factors suitable for integration of safety hazards, security risks, and human factors concerns. To address this need, the aim of this PhD research is focused towards identifying challenges for assessing security-by-design processes, human factors techniques and safety analysis methods. The process-techniques are aligned for validating concerns with the involvement of tool-support. As the use of tool-support minimises and automates the effort of safety, security and human factors experts.

In the following sub-sections, the research findings are summarised towards understanding the integration between safe, secure and usable design systems. The two key findings are: bridging safety, security and human factors, and implementation of design framework.

7.2.1 Bridging Safety, Security and Human Factors

The literature review in Chapter 2 acts as the source of evidence behind the integration of safety, security and human factors engineering. At its core *risk* is identified as an intersecting concept between safety and security. Here, the security challenge is categorised as malicious risk and safety hazard is defined as accidental risk. However, the malicious risk may have safety implications as well, such that safety and security becomes dependent and related to each other.

Usually security breaches are due to risks and hazards which cause accidents. Security and safety have mutual attributes as well such as, dependability defined as justifiably trusted services. The dependability in safety comes along with availability, reliability, integrity and maintainability, whereas in security it comes along with availability, integrity and confidentiality. Therefore, risk factors along with trust and reliance on system are triggered by safety as well as security issues.

Another aspect is the involvement of human factors in the form of human error sources 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), 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, thereby leading to a risk instigating a safety hazard.

Based on this bridging of gap between safety, security and human factors engineers the process-techniques and tool-support options are combined together. Already the safety and security engineers are using many approaches such as, asset modelling, vulnerability and threat identification, risk and hazard analysis. Here, the IRIS framework is used for security-by-design approaches and STPA as hazard analysis method. Furthermore, this contributes towards the identification of task and goal-obstacle modelling as the human factors approach which leads to better safety and security design analysis for critical infrastructures like rail. The design specifications based on use-case template for TA has evidence for highlighting human factors issues. The human factors issues are recognised as the human error sources using HFACS framework. These issues are further categorised using CTA for determining workload behind tasks and HTA for hierarchical breakdown of tasks and use-cases. Thus, helping security and safety engineers to work in collaboration with human factors experts.

7.2.2 Implementation of Design Framework

The safe, secure and usable design framework as presented, is implemented in Chapters 4, 5, and 6 for validation purposes. In Chapter 4, the real-life incident of 'Polish Tram' is applied in a qualitative research based case study evaluation for validation of integration

of security with safety and human factors engineering. Based on the IRIS and CAIRIS modelling process-techniques, assumptions about the security concerns, potential safety hazards and human factors issues in the form of human error sources using HFACS are made. The human factors and safety consultants from Ricardo Rail were presented with this design framework during a workshop, and they suggested an implementation and preparation of a technical specification as a part of their live project. This further strengthened the confidence in design framework.

Also, this part of design framework has been published in Springer workshop which adds to its review and feedback. Hence, reinforcing the bond between safety, security and human factors design approaches.

In Chapter 5, the ERTMS role of 'Signaller' is applied in use-case specifications informed TA approach for validation of process-techniques and tool-support options for design framework. Meeting secure and usable design goals needs the combined effort of safety, security and human factors experts. Human factors experts rely on a combination of cognitive and hierarchical task analysis techniques to support their work. An approach is presented where use-case specifications are used to support TA, and human failure levels help identify design challenges leading to errors or mistakes. An illustration of this approach is by prototyping the role of the ERTMS - Signaller, which provides human factors experts a chance to work in collaboration with safety and security design experts. This part of approach is also accepted for publication in Springer and its feedback adds to its validation.

In Chapter 6, the case study of 'Cambrian Incident' is used to evaluate design framework as an exemplar for resolving safe, secure and usable design issues in rail. Using CAIRIS, the hazard analysis method of STPA is implemented which is derived from security and human factors approach. Eventually, this helps to understand the integration of concepts along with processes and tools. This integration enables the experts to work in collaboration with each other for achieving better design goals in rail infrastructure. A research paper from this approach has been submitted for publication.

Therefore, security risks and human factors issues discovered during IRIS framework and open source CAIRIS tool-support are used to inform potential safety hazards. Usually, these identified safety hazards are mitigated after conducting STPA. However, there is no known software for application of these processes and techniques. Using this design framework, the IRIS framework and CAIRIS tool-support are used to conduct use-case specifications informed TA and STPA based on an integration of safety, security and human factor engineering approaches. This integration intends to support grounds for design framework which is an exemplar for resolving safety, security and human factors issues for critical infrastructures.

7.3 Research Challenges and Limitations

During this PhD research and validation numerous challenges and limitations were faced. For example, during literature review and survey the biggest challenge was identifying the initial point of research. Due to vastness of safety, security, and human factors engineering domains the scope of overlap with regard to research gap was tricky. For this purpose, each domain was thoroughly studied in parallel with discovery of converging concepts. On the basis of this intersection the process-techniques and tool-support options were explored. Also, each concept was backed-up by strong literature evidence.

Another challenge was during Case Study I, where the approach recognition and application was happening, simultaneously. This was due to complex nature of concepts which were explored while working on approach for case study. Here, along with literature survey the consultation from field experts during application of case study was helpful. Using their feedback and review a lot of issues were timely resolved. Similarly, dealing with a wide range of experts from safety, security, and human factors engineering during this research was another challenge. As each expert has a certain knowledge base based on difference of opinion. Here, the target was to bring them all together where each stakeholder's input and feedback was valuable and essential for research. Meanwhile, working on two research papers for publication and writing of *Major Review* for submission was also in progress.

Another limitation faced was during application of Case Study II, where the availability of selected participants for semi-structured interviews i.e. ex-signalman from *Network Rail* was affected due to COVID-19 pandemic. Due to the pandemic, their staff members were working from home. Despite of working at maximum capacity, they were short staffed because of losing people in sickness and facing organisational changes. Things were just beginning to get back to normal work load for office staff but the operational teams were still at full stretch. Therefore, Network Rail staff members can not be contacted for unforeseeable future. In order to overcome this limitation, new plan for short-listing of ex-signalman was done.

Also, at this stage the implementation in CAIRIS was in progress while writing-up of use-case specifications informed TA approach for publication was also due. Here, the biggest challenge was to understand and read the code-work of CAIRIS in terms of database structure and stored procedures. Also, the *Python* programming language skills were polished for scripting while keeping in mind the time constraint.

Finally, during application of Case Study III the data and process validation was to be done by safety and security experts with expertise of human factors processes and techniques. Ideally, the experts from *Ricardo Rail* who were the major beneficiaries of this research were supposed to implement this design framework in their live project. But due to time constraints and other issues, the concept of empirical evaluation was

suggested for this research. Therefore, the experts were consulted for validation of data and process behind design framework using feedback.

7.4 Future Work

7.4.1 Application of Safe, Secure and Usable Design Framework

As future work, the application of safe, secure and usable design framework by different categories of stakeholders from critical infrastructures is suggested, for further feedback and validation. There is scope of improvement after applying this framework for various critical infrastructures. These results and findings will help to identify the factors for improving framework and will also expand the application beyond different infrastructures. Thus, helping experts from various domains by equipping them with tools for achieving better safety, security and human factors design goals.

At present, there has been an on-going research at Massachusetts Institute of Technology (MIT), USA using STPA model to identify human cognitive processes, as means of fleshing out unwarranted assumptions. At a meeting with one their researchers, the facilitation of STPA process-model using integrated design framework was explained to them. They have responded positively towards the approach because it has an empirical basis for model generation (via personas/ argumentation models). As they are not just mere assumptions about the STPA control structures. Hence, there are chances they will be willing to take on with the approach involving process-techniques and tool-support for design framework.

7.4.2 Industrial Viewpoint

Furthermore, there is scope for more process-techniques and tools from safety, security and human factors engineering to be integrated together. Using an industrial viewpoint will help to point out concerns regarding usability of tools. This usability will enhance the tool for resolving design issues. Here the design requirements of tool can be evaluated and validated using an industrial perspective regarding latest trends. These trends will lead to useful insights from an empirical evaluation perspective and add more value to design framework.

One such application is suggested by Ricardo Rail, who are the major beneficiaries of this research project. The design framework has been presented to them at 'Bombardier Aventura - Human Factors Workshop', where the consultants has shown deep interest towards applying this design framework for their live projects. According to them, Ricardo-Roke team will produce a technical specification for their risk analysis project, by involving the human factors work from Aventura. The results will be shared after evaluating their policies around confidentiality of data.

References

2008. RISI - The Repository of Industrial Security Incidents. <https://www.risidata.com/>.
2009. An Arduino universal remote: Record and playback IR signals.
2014. Randstad Rail - Generic Risk Assessment Log.
2017. 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>.
2019. 'Lessons learnt' over train speeding on Cambrian line. *BBC News*.
- Abdulkhaleq, A., Vost, S., Wagner, S. and Thomas, J., 2016. An Industrial Case Study on the Evaluation of a Safety Engineering Approach for Software-Intensive Systems in the Automotive Domain, 27.
- Adelard, 2019. Adelard Overview and Capabilities. <https://www.adelard.com/asce/choosing-asce/index/>.
- Affairs, A. S. f. P., 2013. Task Analysis. </how-to-and-tools/methods/task-analysis.html>.
- Al-Shargie, F., Tariq, U., Mir, H., Alawar, H., Babiloni, F. and Al-Nashash, H., 2019. Vigilance Decrement and Enhancement Techniques: A Review. *Brain Sciences*, 9 (8).
- Alper, S. J. and Karsh, B.-T., 2009. A systematic review of safety violations in industry. *Accident Analysis & Prevention*, 41 (4), 739–754.
- Altaf, A., Faily, S., Dogan, H., Mylonas, A. and Thron, E., 2019. Identifying Safety and Human Factors Issues in Rail using IRIS and CAIRIS. *CyberICPS 2019: 5th Workshop On The Security Of Industrial Control Systems & Of Cyber-Physical Systems*, Luxembourg, Luxembourg: Springer.
- Altaf, A., Faily, S., Dogan, H., Mylonas, A. and Thron, E., 2021. Use-Case Informed Task Analysis for Secure and Usable Design Solutions in Rail. *The 16th International Conference on Critical Information Infrastructures Security*, Lausanne, Switzerland: Springer.

- Anna, G., 2019. Methodological Findings from Applying STPA in Cyber Security Case Studies. *MIT STAMP Conference*, MIT Campus, Cambridge, USA: MIT Partnership for Systems Approaches to Safety and Security (PSASS).
- Atzeni, A., Cameroni, C., Faily, S., Lyle, J. and Flechais, I., 2011. Here's Johnny: A Methodology for Developing Attacker Personas. *2011 Sixth International Conference on Availability, Reliability and Security*, Vienna, Austria: IEEE, 722–727.
- Avizienis, A., Laprie, J.-C., Randell, B. and Landwehr, C., 2004. Basic concepts and taxonomy of dependable and secure computing. *IEEE Transactions on Dependable and Secure Computing*, 1 (1), 11–33.
- Babeshko, E., Kharchenko, V. and Gorbenko, A., 2008. 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*.
- Baysari, M. T., McIntosh, A. S. and Wilson, J. R., 2008. Understanding the human factors contribution to railway accidents and incidents in Australia. *Accident Analysis and Prevention*, 40 (5), 1750–1757.
- Bloomfield, R., Bendele, M., Bishop, P., Stroud, R. and Tonks, S., 2016. The Risk Assessment of ERTMS-Based Railway Systems from a Cyber Security Perspective: Methodology and Lessons Learned. T. Lecomte, R. Pinger and A. Romanovsky, eds., *Reliability, Safety, and Security of Railway Systems. Modelling, Analysis, Verification, and Certification*, Cham: Springer International Publishing, volume 9707, 3–19.
- Bloomfield, R., Bishop, P., Butler, E. and Stroud, R., 2018. Security-Informed Safety: Supporting Stakeholders with Codes of Practice. *Computer*, 51 (8), 60–65.
- Brazendale, J., 1995. IEC 1508: Functional Safety: Safety-Related Systems. *Proceedings of Software Engineering Standards Symposium*, 8–17.
- Brostoff, S. and Sasse, A., 2001. Safe and Sound: A Safety-Critical Approach to Security, 10.
- Cacciabue, P., 2005. Human error risk management methodology for safety audit of a large railway organisation. *Applied Ergonomics*, 36 (6), 709–718.
- Cao, S. and Liu, Y., 2015. Modelling workload in cognitive and concurrent tasks with time stress using an integrated cognitive architecture. *International Journal of Human Factors Modelling and Simulation*, 5, 113.
- CAPEC, 2007. Common Attack Pattern Enumeration and Classification.

- Chen, Y. and Cheng, M., 2015. How Do Designers Deal With Uncertainty, 92.
- Cockburn, A., 26-October- 1998. Basic Use Case Template. (2), 8.
- Cockburn, A. and Bank, N., 1997. Structuring Use cases with goals.
- Conway, D., Dick, I., Li, Z., Wang, Y. and Chen, F., 2013. The Effect of Stress on Cognitive Load Measurement. P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson and M. Winckler, eds., *Human-Computer Interaction – INTERACT 2013*, Berlin, Heidelberg: Springer, Lecture Notes in Computer Science, 659–666.
- Cooper, A., 1999. *The Inmates Are Running the Asylum*. Macmillan Publishing Co.
- Corporation, M., 2015. MITRE ATT&CK®. <https://attack.mitre.org/>.
- Crandall, B., Klein, G. and Hoffman, R. R., 2006. *Working Minds: A Practitioner's Guide to Cognitive Task Analysis*. Cambridge, MA: MIT Press.
- Dardenne, A., van Lamsweerde, A. and Fickas, S., 1993. Goal-directed requirements acquisition. *Science of Computer Programming*, 20 (1), 3–50.
- Davis, W. and Burton, A., 1991. Ecological Task Analysis: Translating Movement Behavior Theory into Practice. *Adapted Physical Activity Quarterly*, 8, 154–177.
- Diaper, D. and Stanton, N., 2004. *The Handbook of Task Analysis for Human-Computer Interaction*. Mahwah, NJ: CRC Press.
- Embrey, D., 2000. Task Analysis Techniques, 14.
- Embrey, D. D. and Zaed, S., 2021. A Set Of Computer Based Tools Identifying And Preventing Human Error In Plant Operations, 11.
- Erbacher, R. F., Frincke, D. A., Wong, P. C., Moody, S. and Fink, G., 2012. A Multi-Phase Network Situational Awareness Cognitive Task Analysis:. *Information Visualization*.
- European Network and Information Security Agency, 2020. *Railway Cybersecurity: Security Measures in the Railway Transport Sector*. LU: Publications Office.
- Faily, S., 2018. *Designing Usable and Secure Software with IRIS and CAIRIS*. Cham: Springer International Publishing.
- Faily, S. and Fléchais, I., 2010. Barry is not the weakest link: Eliciting Secure System Requirements with Personas, 8.
- Faily, S. and Flechais, I., 2011. User-Centered Information Security Policy Development in a Post-Stuxnet World. *2011 Sixth International Conference on Availability, Reliability and Security*, Vienna, Austria: IEEE, 716–721.

- Faily, S. and Fléchais, I., 2016. Finding and resolving security misusability with misusability cases. *Requirements Engineering*, 21 (2), 209–223.
- Felice, F. D. and Petrillo, A., 2011. Methodological Approach for Performing Human Reliability and Error Analysis in Railway Transportation System, 13.
- Folse, S. A., 2017. Systems-Theoretic Process Analysis of Small Unmanned Aerial System Use at Edwards Air Force Base, 128.
- France, M. E., 2017. *Engineering for Humans : A New Extension to STPA*. Thesis, Massachusetts Institute of Technology.
- Golightly, D., Balfe, N., Sharples, S. and Lowe, E., 2009. Measuring situation awareness in rail signaling. *Rail Human Factors Around the World: Impacts on and of People for Successful Rail Operations*, 361–369.
- Gollmann, D., 2007. *Computer Security*. Wiley & Sons, second edition.
- Grivicic, D., 2019. *Can STAMP Provide a Complete Safety Argument?*. Ph.D. thesis.
- HA, 2019. Hazard Analysis — Risk Assessment — Safety System — IT Risk Management. http://www.chambers.com.au/glossary/hazard_analysis.php.
- Hammerl, M. and Vanderhaegen, F., 2009. Human Factors In The Railway System Safety Analysis Process: 3rd International Rail Human Factors Conference, 9.
- Harrison, H., Birks, M., Franklin, R. and Mills, J., 2017. Case Study Research: Foundations and Methodological Orientations. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 18 (1).
- IEC, 2019. IEC - TC 65/AHG 1 Dashboard > Structure: Subcommittee(s) and/or Working Group(s), Membership, Officers, Liaisons. https://www.iec.ch/dyn/www/f?p=103:14:0:::FSP_ORG_ID,FSP_LANG_ID:11734,25.
- ISO/IEC, 2004. ISO/IEC 13335-1:2004. <http://www.iso.org/>.
- ISO/IEC, 2007. ISO/IEC 27002: Information Technology – Security Techniques – Code of Practice for Information Security Management.
- ISO/IEC, 2010. ISO 9241-210:2010. <http://www.iso.org/>.
- Jen, R., 2012. How to increase risk awareness. *PMI® Global Congress 2012*, Vancouver, British Columbia, Canada, North America: PA: Project Management Institute.
- Jones, A. and Ashenden, D., 2005. Risk management for computer security : Protecting your network and information asset. Butterworth-Heinemann.

- Jonsson, E. and Olovsson, T., 1998. On the Integration of Security and Dependability in Computer Systems, 6.
- Karanikas, N., 2016. Human Factors Science and Safety Engineering. Can the STAMP Model Serve in Establishing a Common Language? 132–149.
- Karatzas, S. and Chassiakos, A., 2020. System-Theoretic Process Analysis (STPA) for Hazard Analysis in Complex Systems: The Case of “Demand-Side Management in a Smart Grid”. *Systems*, 8 (3), 33.
- Kirwan, B., 1997. Validation of human reliability assessment techniques: Part 1 — Validation issues. *Safety Science*, 27 (1), 25–41.
- Kriaa, S., Pietre-Cambacedes, L., Bouissou, M. and Halgand, Y., 2015. A survey of approaches combining safety and security for industrial control systems. *Reliability Engineering & System Safety*, 139, 156–178.
- Lahoz, C. H. N., 2015. Systematic review on STPA A preliminary study, 34.
- Leveson, N., 2011. Engineering a Safer and More Secure World, 72.
- Leveson, N., 2018. Systems-Theoretic Process Analysis (STPA) - Handbook, 188.
- Leveson, N. G., 1987. Software Fault Tolerance in Safety-Critical Applications. F. Belli and W. Görke, eds., *Fehlertolerierende Rechensysteme / Fault-Tolerant Computing Systems*, Springer Berlin Heidelberg, Informatik-Fachberichte, 1–12.
- Mastery, U., 2012. UX Techniques -. <https://uxmastery.com/resources/techniques/>.
- Maxion, R. A., Longstaff, T. A. and McHugh, J., 2010. Why is There No Science in Cyber Science?: A Panel Discussion at NSPW 2010. *Proceedings of the 2010 New Security Paradigms Workshop*, New York, NY, USA: ACM, NSPW '10, 1–6.
- Mellado, D., Fernández-Medina, E. and Piattini, M., 2006. Applying a Security Requirements Engineering Process. D. Gollmann, J. Meier and A. Sabelfeld, eds., *Computer Security – ESORICS 2006*, Berlin, Heidelberg: Springer, Lecture Notes in Computer Science, 192–206.
- Militello, L. and Hutton, R., 1998. Applied Cognitive Task Analysis (ACTA): A Practitioner’s Toolkit for Understanding Cognitive Task Demands. *Ergonomics*, 41, 1618–41.
- Mindermann, K., Riedel, F., Abdulkhaleq, A., Stach, C. and Wagner, S., 2017. Exploratory Study of the Privacy Extension for System Theoretic Process Analysis (STPA-Priv) to Elicit Privacy Risks in eHealth. *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*, Lisbon, Portugal: IEEE, 90–96.

- Nielsen, L., 2013. *Personas - User Focused Design*. Human–Computer Interaction Series, London: Springer-Verlag.
- 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., Kourssi, E. M. E. and Ghazel, M., 2016. The New Challenges of Rail Security. *Journal of Traffic and Logistics Engineering*.
- Norman, D., 2004. *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books.
- O'Hare, D., 2001. *The 'Wheel of Misfortune': A Taxonomic Approach to Human Factors in Accident Investigation and Analysis in Aviation and Other Complex Systems*, volume 43.
- O'Riordan, D., 2015. Safety integrity levels and system design risk analysis.
- Ouferroukh, H., 2018. Safety Culture. https://www.era.europa.eu/activities/safety-culture_en.
- OWASP, 2001. The Open Web Application Security Project.
- Papa, M. and Sheno, S., 2008. *Critical Infrastructure Protection II*. Springer Science & Business Media.
- Peper, N. A., 2017. Systems Thinking Applied to Automation and Workplace Safety, 109.
- Pereira, D., Hirata, C., Pagliares, R. and Nadjm-Tehrani, S., 2017. Towards Combined Safety and Security Constraints Analysis. S. Tonetta, E. Schoitsch and F. Bitsch, eds., *Computer Safety, Reliability, and Security*, Cham: Springer International Publishing, volume 10489, 70–80.
- Pereira, D. P., Hirata, C. and Nadjm-Tehrani, S., 2019. A STAMP-based ontology approach to support safety and security analyses. *Journal of Information Security and Applications*, 47, 302–319.
- Peters, J., 2019. MITRE ATT&CK Framework: Everything You Need to Know. <https://www.varonis.com/blog/mitre-attck-framework-complete-guide/>.
- Piètre-Cambacédès, L. and Bouissou, M., 2013. Cross-fertilization between safety and security engineering. *Reliability Engineering & System Safety*, 110, 110–126.
- Preece, J., Rogers, Y. and Sharp, H., 2002. What is User Centered Design? <https://www.interaction-design.org/literature/topics/user-centered-design>.

- Pruitt, J. and Grudin, J., 2003. Personas: Practice and Theory. *Proceedings of the 2003 Conference on Designing for User Experiences*, New York, NY, USA: ACM, DUX '03, 1–15.
- qonita, 2018. Exploratory Design Research Interview. <https://medium.com/designstrat/exploratory-design-research-interview-dc51398c6354>.
- RAIB, 2019. 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>.
- Raspotnig, C. and Opdahl, A. L., 2012. Improving Security and Safety Modelling with Failure Sequence Diagrams. *International Journal of Secure Software Engineering (IJSSE)*, 3 (1), 20–36.
- Rausand, M., 2014. Reliability of Safety-Critical Systems: Theory and Application. *Reliability of Safety-Critical Systems*, John Wiley & Sons, Ltd, i–xviii.
- Rearick, M. L. and Feldman, A., 1999. A Framework for Understanding Action Research, 32.
- Reason, J., 1990. Human Error by James Reason.
- RSSB, 2019. RSSB Tools & Resources. <https://www.rssb.co.uk/en/Standards-and-Safety/Tools-Resources/Human-Factors-Toolkit>.
- Scarinci, A., 2017. MONITORING SAFETY DURING AIRLINE OPERATIONS: A SYSTEMS APPROACH, 83.
- Schneier, B., 1999. Academic: Attack Trees - Schneier on Security by Dr. Dobb's Journal. https://www.schneier.com/academic/archives/1999/12/attack_trees.html.
- Schneier, B., 2000. Secrets and Lies: Digital Security in a Networked World. John Wiley & Sons.
- Sharp, H., Rogers, Y. and Preece, J., 2007. *Interaction Design by Rogers, Yvonne, Preece, Jenny, 1949-, Sharp, Helen*. John Wiley & Sons, 2nd edition edition.
- Shorrock, S. T., 2007. Errors of perception in air traffic control. *Safety Science*, 45 (8), 890–904.
- Shostack, A., 2014. *Threat Modeling: Designing for Security*. Indianapolis, IN: John Wiley and Sons. Adam Shostack.
- Sindre, G. and Opdahl, A. L., 2005. Eliciting security requirements with misuse cases. *Requirements Engineering*, 10 (1), 34–44.

- Slominski, H. M., 2020. Using STPA and CAST to Design for Serviceability and Diagnostics, 106.
- Soegaard, M. and Dam, R. F., 2013. *The Encyclopedia of Human-Computer Interaction, 2nd Ed.*
- Steeves, V., 2018. *NDSR Art Immersion Week: Research with Interviews.*
- Storey, N., 1996. *Safety-Critical Computer Systems.* Harlow: Addison-Wesley. Neil Storey.
- Summers, S. E., 2018. Systems Theoretic Process Analysis Applied to Air Force Acquisition Technical Requirements Development, 184.
- Tech, R., 2017. WannaCry virus was 'wake-up call' for railway industry — RailTech.com. <https://www.railtech.com/digitalisation/2017/12/11/wannacry-virus-was-wake-up-call-for-railway-industry/>.
- UIC, 2019. UIC - Human Factors and Safety Resources. <https://uic.org/safety/human-factors/>.
- W. Edgar, T. and O. Manz, D., 2017. *Research Methods for Cyber Security — ScienceDirect.* Copyright © 2017 Elsevier Inc. All rights reserved.
- Wiegmann, D. A. and Shappell, S. A., 2003. *A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System.* Aldershot, Hants, England ; Burlington, VT: Routledge, 1 edition edition.
- Wiley, J., 2013. *Handbook of Loss Prevention Engineering, 2 Volume Set — Wiley.*
- Winther, R., Johnsen, O.-A. and Gran, B. A., 2001. Security Assessments of Safety Critical Systems Using HAZOPs. *SAFECOMP.*
- Young, W. and Leveson, N. G., 2014. An integrated approach to safety and security based on systems theory. *Communications of the ACM*, 57 (2), 31–35.
- Zhou, J.-L. and Lei, Y., 2018. Paths between latent and active errors: Analysis of 407 railway accidents/incidents' causes in China. *Safety Science*, 110, 47–58.

Appendices

Appendix A

Ethics Approval

During this PhD research, the ethics approval was obtained from BU ethics community as shown in Table A.1. There was a need to understand the major routine tasks performed by a 'Train Signaller', in order to identify the human error sources during interaction with system. Also, the aim was to determine the task breakdown and cognitive attributes responsible for affecting human performance.

The criteria for the choice of participant was someone with work expertise as a train signaller, employed with minimum of two years of experience, and has good understanding of English language. No audio/ video was recorded during interview. The semi-structured interviews were like open conversations, in which participants were encouraged to talk about their work routine. The written notes produced during these interviews were used for research.

Table A.1: Research Ethics Checklist

| | |
|------------------------------|---|
| Ethics ID | 27657 |
| Status | Approved |
| Date Approved | 24/09/2019 |
| Risk | Low |
| Course | Postgraduate Research - Faculty of Science and Technology |
| Project Title | Integrating Safety, Security and Human Factors Engineering in Rail Infrastructure Design and Evaluation |
| RED ID | 143869 |
| External Funding Body | Ricardo Rail |
| Start Date of Project | 17/09/2018 |
| End Date of Project | 17/09/2021 |
| Supervisor | Dr Shamal Faily |
| Approver | Professor Marcin Budka |

Appendix B

Interview Process & Template

During interview process, the participants were asked to provide some basic details like name, email etc. These details allowed to contact them for research purposes, and send them important information regarding research project. The participants were asked to sign the 'Participant Agreement Form'. After that participants were invited for interview. Each interview was 30-45 minutes long in duration.

| | |
|--------------------------------------|--|
| Task Name: | |
| Major Actor/s: | |
| Description: | |
| Pre-Condition/s (Sub-Task/s): | |
| Action Sequence: | Exception/s: |
| Post Condition/s: | |
| Cognitive Attributes: | <ol style="list-style-type: none">1. Vigilance2. Situational Awareness3. Work Load4. Stress5. Risk Awareness |
| Remarks: | |

Figure B.1: Template of Task Sheet for Interview

During interview, participants were asked questions about their working hours and task routine. This included a brief about the major tasks performed by them as a 'signaller'. The participants were asked about specific tasks and sub-tasks, along with the sequence of actions required to complete those tasks. They were also asked about the cognitive attributes responsible for affecting their job performance. A task sheet was

maintained and filled during interview against task/s mentioned. The template of task sheet is shown in Fig. B.1.

Also, all personal data collected for the purposes of this study will be held for the duration of this project which is up-till September 2021. The published research outputs are anonymised.

Appendix C

Case Studies - CAIRIS Model Files

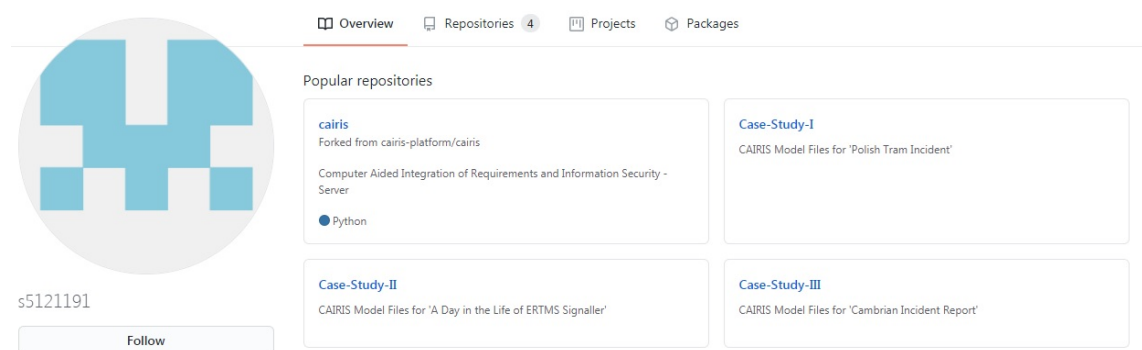


Figure C.1: GitHub repository for CAIRIS Models

When conducting case study research with tool-support from CAIRIS, a number of models were generated. The model files were exported and saved as xml files. These xml files are available at GitHub repository of *s5121191*. The repository contains xml files and other informational sheets (spreadsheets) for case studies in folders named as:

1. **Case Study I:** For CAIRIS model files of 'Polish Tram Incident'.
2. **Case Study II:** For CAIRIS model files of 'A Day in the Life of ERTMS Signaller'. Also, found here is the spreadsheet for use-case specifications.
3. **Case Study III:** For CAIRIS model files of 'Cambrian Incident Report'.

These xml files can be imported into demo version of CAIRIS found online at the link: <https://demo.cairis.org>.