ISSN 2754-1118
Vol.2 No.2



The Safety-Critical Systems Club

SAFETY-CRITICAL SYSTEMS eJOURNAL

SCSC

Editorial to the 2023 Summer Issue

Welcome to the second issue of the second volume of the Safety-Critical Systems eJournal, which is published by the Safety-Critical Systems Club (SCSC). This issue contains three papers:

- Rob Ashmore, Mark Hadley and James Sharp (UK), address "*Reducing the Risk of a Software Common Mode Failure*". They provide examples showing that this is not just a theoretical risk and review some previous research, extant standards, and how the problem has been addressed in real-world systems. They conclude that there is no preferred way of protecting against common mode failure in software-based systems and therefore propose a set of criteria that can be used to assess the protections that have been implemented within a system design.
- Derek Fowler and Alasdair Graebner (UK) build upon Derek's paper in the last issue of Volume 1, on using IEC61508 in the Transport Sector, by providing another worked example, "*An IEC 61508 Viewpoint on the Safety Assessment of Railway Control Systems*". The example considers moving-block Automatic Train Control for a hypothetical underground railway system.
- Malcolm Jones (UK) is "*Chasing the Black Swan*". There may be an 'as yet' undiscovered flaw or lack of understanding in the design of a product, process or facility that could lead to a catastrophic event. How should one continue to search for such a possible flaw with a view to subsequent removal or mitigation — when is 'enough-enough'?

My thanks go to the authors for contributing their papers, and also to the anonymous peer-reviewers (at least three per paper) for suggesting improvements. Apologies also to those reviewers who made some recommendations that were not taken up.

No letters have been received for publication in this issue, but there has been correspondence. An author (not in this issue) and a couple of reviewers independently said that they thought that the referencing of standards, in particular IEC61508, would be clearer were we to follow the recommendations of the standardisation body, rather than the Harvard method we have been using. This issue contains two papers that reference IEC61508, and they do it differently. Derek's paper is the third in a series and so has been kept consistent with its predecessors. I will do a review of referencing and update the Author Guidelines (to be found in the paper template) accordingly.

Previous editorials have said that the next volume of the Safety-Critical Systems eJournal will start with a themed issue. This has been put back a month so that the themed issue will be the second of the year, published to coincide with the Safety-Critical Systems Symposium, SSS'24. The theme is the technologies underpinning autonomous vehicles and how we assure them. Note that this theme is broad — for example "technologies", as well as vision processing, machine learning, etc., can include things like concepts of operation, regulation, standards, ownership of liability, and so on.

Finally a reminder that this year's cover image is to highlight the Systems Approach to Safety of the Environment Working Group, which aims to produce clear guidance on how engineered systems should be developed and managed throughout their entire lifecycle so as to preserve, protect and enhance the environment. If you would like to join, or find out more about this group, please go to their page on the SCSC website: https://scsc.uk/ge

John Spriggs, SCSC Journal Editor
August 2023

ISSN 2754-1118 (Online)
ISSN 2753-6599 (Print)

Reducing the Risk of a Software Common Mode Failure

Rob Ashmore, Mark Hadley and James Sharp

Dstl, Portsdown West, UK

Abstract

We describe the nature of a common mode failure, illustrating why software-based components can be particularly susceptible. A number of historical accidents and incidents are used to demonstrate this is not just a theoretical risk. Previous academic research, existing standards and historical civil air programmes are surveyed. This shows the importance of software common mode failure is recognised, but there is no preferred way of protecting against it. A set of criteria are proposed, which provide a means of assessing protections that have been implemented within a system design. These are used to highlight approaches to protection that are suggested for future air systems (and other safety-related applications).

1 Introduction

In an attempt to increase reliability, system architectures often include replicated items. The desired increase in reliability is based on an assumption, typically implicit, that each item will fail independently. A Common Mode Failure (CMF) is an event that undermines this desired independence.

If the replicated items are running software, then this has the potential to cause a CMF. This potential is heightened by the nature of software: using identical copies of the same software, and providing them with identical inputs, offers no protection against software CMF. This is very different to the situation for mechanical or discrete electronics components where identical items, which fail independently, can provide protection.

We are primarily concerned with protecting against software CMF in future air systems, but note that many of the same considerations apply to other safety-related applications.

Initially, in Section 2, a selection of historical examples illustrate the real-world importance of software CMF. An informal review of key academic literature is provided in Section 3. Treatment of software CMF in a selection of standards (and standards-like) documents is discussed in Section 4. Previous approaches adopted within the civil air domain are briefly described in Section 5. Criteria to assist in evaluating protections against software CMF are listed in Section 6, and are related to potential protection approaches that could be adopted in system designs in Section 7. Closing remarks are provided in Section 8.

© Crown copyright (2023), Dstl. This material is licensed under the terms of the Open Government Licence except where otherwise stated. To view this licence, visit https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3 or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gov.uk

2 Historical Examples

Software CMF is not just a theoretical concern. It has been the cause of a number of accidents and incidents, a selection of which are briefly summarised below:

Ariane 5 (Lions 1996): Two SRIs (Inertial Reference Systems[1]) were operating in parallel, with identical hardware and software. A numeric overflow was detected and, by design, the active SRI shut down. The backup SRI had shut down in the preceding time period for exactly the same reason. The lack of a functioning SRI led the loss of the launcher.

NATS System Failure (Walmsley et al. 2015): NATS run a primary System Flight Server (SFS), with a secondary SFS always ready in order to preserve availability. Prior to this incident, recent changes had exposed a latent fault in the SFS software. When an action caused this fault to become a failure, by design the primary SFS shut down. Since it ran identical software, the secondary SFS also shut down. Up to 1900 flights and 230,000 passengers were affected by the resulting outage.

Subsonic / Supersonic Boundary (JSSSC 2010): A front-line aircraft had been rigorously developed, thoroughly tested and dozens were in operational use. As part of an upgrade, a weapons test required extensive telemetry. Analysis showed that all aircraft computers failed and then restarted as the aircraft passed through Mach 1. The aircraft had sufficient momentum and mechanical control that it coasted through this anomaly without the pilot noticing.

Airbus A330-302 (TTSB 2021): On June 14, 2020, China Airlines flight CI202, an Airbus A330-302 aircraft, took off from Shanghai bound for Taipei. At touchdown, the aircraft experienced the quasi-simultaneous failure of the 3 Flight Control Primary Computers (FCPCs), which meant that ground spoilers, thrust reversers and autobrake became unavailable. The flight crew rapidly applied full manual brake to stop the aircraft safely about 30 feet before the end of the runway. The software at the core of this incident had supported more than 8.7 million flight cycles, with no other observations of this type of issue.

These examples show that software CMF can be a significant issue, even in systems that have been rigorously developed and that have a significant amount of service history.

3 Academic Research

3.1 Fault Avoidance

Avoiding the introduction of faults, as far as possible, is the first way of protecting against software CMF. For example, Powell et al. (2011) note that: "*The first defense against design faults is of course to attempt to avoid their occurrence in the first place by applying the techniques of fault prevention (i.e., 'careful design') and fault removal (i.e. 'verification and testing')*".

(Note that, generally speaking, the academic literature uses the term "design fault" to include faults from a number of software development phases, including, requirements, design, implementation, and maintenance.)

[1] SRI is Système de Référence Inertiel, i.e. Inertial Reference System

Fault avoidance typically involves the application of appropriately rigorous software development and assurance standards; for example, RTCA/DO-178C (RTCA 2011).

3.2 Fault Tolerance

Powell et al. (2011) also discuss protection against software CMF under the general heading of tolerance to design faults. The first-level response to tolerating design faults comes in two forms, data diversity and design diversity. Both of these forms aim to prevent identical software being presented with identical inputs, thus introducing some (typically, limited) independence.

Data diversity is focused on inputs to the software, whereas design diversity is concerned with the software itself. Broadly speaking there are three basic approaches (see, for example, Powell et al. (2011), Siewiorek et al. (2004) and Laprie et al. (1990)) to introducing design diversity, these being recovery blocks, n-version programming, and n-self checking.

The following subsections provide further information on approaches to fault tolerance.

3.3 Data Diversity

This type of diversity can be achieved by providing identical software with different inputs, for example from sensors located at different points on the airframe. The key premise behind data diversity is that software problems arising during operational use are subtle and hence only triggered for small parts of the input space (see, for example, Ammann and Knight (1988)). Hecht (1993) collected evidence from a number of sources, including spacecraft avionics and telephone switches, which also supports this view.

However, only a limited amount of diversity can be introduced by providing identical software with different input data. In particular, this data typically comes from different sensors that are sampling the same (or at least a very similar) real-world situation. It is unlikely, for example, that two properly-functioning air data sensors will provide radically different data.

One situation in which wildly different data could be provided is following a sensor failure. The software would be expected to handle this failure, but data diversity, in the form of different sensors providing inputs to different copies of the software, would provide an additional layer of protection. An alternative way of providing protection against this specific type of problem would be checking inputs before they are used and, if necessary, replacing out-of-range values with a known-safe alternative.

In theory, it is possible to create artificial data diversity, for example, by adding noise to input signals. In practice it can be difficult to determine an appropriate way to alter inputs, so that their essential meaning is preserved but they are sufficiently different to provide some protection against software CMF.

3.4 Design Diversity — General

The intent of design diversity is to create versions of software that fail in non-identical ways. This typically involves changes to earlier parts of the software development lifecycle (e.g. design and implementation).

Design diversity typically provides greater protection than data diversity as, for example, it is unlikely that two separate programming teams will have made identical errors. That said, as discussed below, common mode failures can still occur even with non-identical

software. Furthermore, design diversity incurs greater costs, as additional software development is required.

3.5 Design Diversity — Recovery Blocks

Randell introduced the concept of recovery blocks in 1975 (Randell 1975). These were motivated by the observation that it is standard practice to structure a program of any significant complexity into a series of blocks (e.g. modules, procedures, subroutines, methods).

Before a "significant" block is executed, a recovery point is stored. The first alternate block is run and its output is subjected to an acceptance test. If the output is not acceptable then the program is rewound to the recovery point, the second alternate block is executed and its output is tested for acceptance. The process is repeated until either an acceptable output is found or all alternates have been exercised.

Theoretical issues with the recovery block approach include inaccuracies in the acceptance test, and the possibility of interacting tasks causing a domino effect of cascading recoveries. Tomek et al. (1993) also note that it can be difficult to achieve diversity in the alternate implementations of the recovery blocks; this is partly due to their relatively small size, and partly due to the fact that they are forced to operate on identical inputs.

An experimental evaluation of the recovery block approach was conducted by Anderson et al. (1985). This was based on what was then considered to be a medium-scale development (approximately 8000 lines of CORAL[2] 66), specifically, a naval command and control system. The software was developed by professional programmers to "normal commercial" standards. Of 222 potential failures identified during the evaluation, 165 (approx. 74%) were adequately addressed using the recovery block mechanism. The authors believe that the prototype recovery algorithms (i.e. those for storing and rewinding to the recovery point) were a limiting factor in their results. The use of recovery blocks did, however, lead to larger code size, extra memory demands and longer run time.

3.6 Design Diversity — N-Version Programming

In this approach, several alternate versions are run in parallel and their output is passed to an adjudicator. The adjudicator provides a consolidated output for use by the rest of the system. For example, this consolidation could check that outputs are within pre-defined bounds and then return an average; another alternative would be the use of a simple majority voter.

One of the earliest evaluations of n-version programming was performed in 1996 by Knight and Leveson. In this study (Knight and Leveson 1996), 27 different students each wrote a program to satisfy the same set of requirements. Each program had to pass a series of acceptance tests. Once each program was "acceptable", the collection of programs was subjected to a large barrage of tests. Several common faults were detected between the programs and, more significantly, the number of common faults was higher than would be predicted if the programs exhibited independent faults.

A more recent study (from 2006) was conducted by van der Meulen and Revilla. This considered over 36,000 programs that were submitted as potential solutions to problems posed on a programming website. They concluded (van der Meulen and Revilla 2006) that

[2] CORAL is the Computer On-line Real-time Applications Language

using different languages led to more diverse programs (e.g. authors were more likely to code a for loop incorrectly when using C than when using Pascal) and, furthermore, that the benefit of multi-version programming reduces as the individual program versions themselves get more reliable.

One problem that n-version programming cannot solve is that of an incorrect requirement. There are also other aspects of programs where apparent diversity could be undermined. For example, Voas et al. (1997) cite a case where common mode failures were injected in program initialisation, despite other aspects of the programs being quite diverse.

A weakness of the three studies referenced above is that they are based on programs written either by students or by "hobby programmers". None of them is based on the type of software that would emerge from a rigorous development process like that associated with aircraft systems. It is not clear, for example, how limits on allowable cyclomatic complexity may constrain implementation options. Likewise, the extent to which software education implicitly reduces diversity is not clear, nor is it clear whether international development programmes offer an opportunity to increase this type of design diversity, for example, due to different educational approaches and cultural norms (Olson and Olson 2003).

The main conclusion to draw from the above examples is that whilst n-version programming can offer some benefits it does not protect against all software common mode faults. Furthermore, the level of benefit that it provides is less than would follow from a naive assumption of complete independence between the versions.

3.7 Design Diversity — N-Self Checking

The n-self checking approach can be considered as a hybrid of the other two approaches. It involves the use of at least two self-checking software components. A component could be made by combining a version of the software and an acceptance test. Alternatively, it could be made by combining several versions of the software with an adjudicator.

The approach of n-self checking software has been studied less extensively in academia than either recovery blocks or n-version programming. Despite that, as indicated later (in Section 5), it forms the basis of at least one system in the civil sector.

3.8 Design Diversity — Different Functionality

In some applications it may be possible to achieve design diversity by implementing software that achieves the same system-level outcome but using different functionality. The use of a different control law in an aircraft flight control system is a potential example. Likewise, implementing software deliberately intended to manage degraded modes of operation, is another.

In some cases, an easy way of achieving some design diversity may be to use an older version of software alongside the most recent one. This might be appropriate if the change between the two software versions was solely related to performance issues. It is not appropriate, however, if the new version fixed a safety-related issue.

3.9 Watchdog Timers

The previous discussions have focussed on fault tolerant approaches that are embodied in software. Another approach involves the use of a watchdog timer: see, for example, Namjoo and McCluskey (1995).

In general, the system is designed so that it signals the watchdog once during each pre-scheduled interval. If the timer is not reset (and, in some cases, if it is reset more than once) during an interval then the watchdog raises an error. This error can be trapped and used, for example, to restart a processor.

Hence, watchdog timers, which are often implemented in hardware, can be used to provide protection against particular types of software error, including those that cause applications to crash and those that result in infinite loops.

3.10 Summary

The academic literature includes a variety of mechanisms for protecting against software design faults that would otherwise cause CMF. Avoiding, as far as possible, the introduction of faults by the application of suitably rigorous software development processes, is important but it does not fully resolve the problem. Additional mechanisms for protecting against software CMF have different strengths and weaknesses. In particular, there is no reliable way of entirely removing the possibility of software CMF: when multiple copies of software are used, it will always represent a risk. Likewise, whilst the value of diversity is noted, there is no preferred way of mitigating this risk. Furthermore, there does not appear to be a standard way of assessing the level of protection (or risk mitigation) provided by a particular approach, or combination of approaches; as indeed is highlighted in the introduction to Annex C, Part 3 of IEC 61508:2010:

> *"Given the large number of factors that affect software systematic capability it is not possible to give an algorithm for combining the techniques and measures that will be correct for any given application."*

4 Standards

The following paragraphs summarise how software CMF is covered in a number of standards, including ones specific to the air domain as well as ones applicable to other domains that use safety-related software.

DEF STAN 00-970: This UK Defence Standard, "*Certification Specifications for Airworthiness*" (UKMOD 2020), does not directly highlight CMF as an issue, whether specific to software or in any other context.

However, software CMF is indirectly included via references that describe an Acceptable Means of Compliance (AMC). For example, Requirement UK25.1309e is concerned with safety-related Programmable Elements (PE). The associated AMC references ARP 4761 (SAE 1996) and RTCA/DO-178C (RTCA 2011). As discussed below, both of these include specific comments on software CMF.

MIL-HDBK-516C: This U.S. Department of Defense Military Handbook, "*Airworthiness Certification Criteria*" (USDoD 2014), includes several mentions of common mode failure.

For example, within Section 6, "Flight Technology", to satisfy one of the criteria requires "*the probability of a common mode failure is extremely remote ($1x10^{-9}$ or as specified by the procuring activity) and is verified by fault tree and hazard analysis*". Likewise, in Section 7, "Propulsion and Propulsion Installations", satisfying one of the criteria requires

"*controls and subsystems are systemically and operationally isolated to avoid possible cascading failures due to any single or common cause*".

These extracts emerge from subsystem considerations, specifically vehicle control and propulsion. Within MIL-HDBK-516C, software-specific considerations are in Section 15, "Computer Systems and Safety". That section notes typical certification source data include "*various technical analyses and studies (e.g. common mode analysis, ...)*".

These extracts show that CMF is highlighted in MIL-HDBK-516C, both in general and specifically in relation to software. However, there is no detailed information on how best to protect against this possibility.

ARP 4761: This Aerospace Recommended Practice, "*Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment*" (SAE 1996), records the potential for CMF, both in general and from the specific perspective of software. It provides a range of techniques for identifying elements that are potentially susceptible to CMF. However, it does not provide a description of ways of protecting against software CMF.

RTCA/DO-178C: This *de facto* assurance standard, "*Software Considerations in Airborne Systems and Equipment Certification*" (RTCA 2011), discusses architectural considerations, with further subsections considering partitioning, multiple-version dissimilar software and safety monitoring.

It notes that partitioning may be achieved by using different hardware for each software component. It also notes that, alternatively, partitioning provisions may be made to allow multiple software components to run on the same hardware platform. Regardless of the adopted approach, partitioned software components should not affect each other's code, inputs, outputs, or data storage.

With respect to multiple-version dissimilar software, RTCA/DO-178C notes that processes completed before dissimilarity is introduced remain a potential source of common errors (e.g. n-version software being developed from the same set of requirements). It also notes that the degree of dissimilarity, and hence the degree of protection, is not usually measurable. As such, dissimilarity is usually used to offer additional protection, beyond that achieved by other DO-178C activities.

IEC 61508: Typically, aviation-specific standards would be preferred to IEC 61508, which adopts a more generic outlook. However, IEC 61508 can still provide useful guidance. For example, in relation to software requirements, Part 3 thereof (IEC 61508-3:2010) notes that:

> *"In order to address independence, a suitable common cause failure analysis shall be carried out. Where credible failure mechanisms are identified, effective defensive measures shall be taken."*

Thus, the importance of protecting against common mode (or, using the terminology of IEC 61058, common cause) failure is also highlighted in relation to both software architecture and software design.

In addition, Annex C thereto includes guidance on a number of techniques and measures that may offer some defence against common mode failure. Examples include fault detection, diverse monitor techniques, re-try fault recovery mechanisms, and graceful degradation.

Nuclear: Whilst it would not be appropriate to apply nuclear standards to aviation directly, they can nevertheless provide informative guidance. One example of this is the US

Nuclear Regulatory Commission (NRC) Regulatory Guide (RG) 1.152, "*Criteria for Use of Computers in Safety Systems of Nuclear Power Plants*" (USNRC 2011), which states:

> *"With the introduction of digital systems into plant safety system designs, concerns have emerged about the possibility that a design error in the software in redundant safety system channels could lead to a common-cause failure or common-mode failure of the safety system function. Conditions may exist under which some form of diversity may be necessary to provide additional assurance beyond that provided by the design and quality assurance programs that incorporate software quality assurance and verification and validation."*
>
> *"With respect to software diversity, experience indicates that the independence of failure modes may not be achieved in cases in which multiple versions of software are developed from the same software requirements."*

Again, this guidance highlights the possibility of software CMF, whilst also noting that some factors (e.g. an incorrect requirement) can defeat assumed levels of software independence.

More-detailed, software-specific guidance for the nuclear industry is provided by a "common position of international nuclear regulators and authorised technical support organisations" on "licensing of safety critical software for nuclear reactors" (TF SCS 2022). Released in 2021 and revised the following year, this recent document notes, for example:

> *"The potential for common cause software failures across all defence barriers utilising computer based systems should be analysed."*
>
> *"Options must be taken as to what are the most effective methods and techniques to ensure diversity of failure behaviour across diverse software programs. ... Current practice is founded on engineering judgment, which identifies what are considered to be the best means to force diversity with some limited support from research. The possibility of providing further scientific support to these judgments is an open research issue."*

These extracts (and other content) highlight the importance of software CMF, whilst recognising there is no preferred way of introducing diversity to protect against it.

Summary: Software CMF is highlighted in a number of standards and guidance documents. In some cases, possible ways of protecting against this are mentioned (including partitioning and n-version programming). There is, however, no consensus on a preferred approach to protecting against software CMF.

5 Civil Air Sector

5.1 Airbus A320

The following description is based on that provided by Brière and Traverse (1993).

The Airbus A320 was the first civil aircraft equipped with a digital electrical flight control system. This system is designed to tolerate both hardware and software design faults. The aircraft was certified and entered service in the first quarter of 1988.

Two dissimilar types of computer are used in the flight control system:

- Elevator and Aileron Computers (ELACs). These are manufactured by Thomson CSF and are based around the 68010 microprocessor.
- Spoiler and Elevator Computers (SECs). These are based on the 80186 microprocessor and are manufactured in conjunction with SFENA/Aerospatiale.

Each computer features a control channel and a monitoring channel. That is, each computer is self-checking (in the sense of being part of an n-self checking implementation). In total, four different pieces of software exist, covering all combinations of type of computer (i.e. ELAC or SEC) and type of channel (i.e. control or monitor).

The software was developed to RTCA/DO-178A Level 1, which was at that time the most stringent civil aviation software standard. The basic rule adopted by Airbus was that the software is built in the best possible way. Dissimilarity between the four pieces of software is an additional precaution; it is not used to reduce the required software quality effort. (This matches the approach espoused in the more recent RTCA/DO-178C. (RTCA 2011))

For any given function, one computer is designated as being active, with the others being available as "hot spares". Failure detection is achieved by comparing the control output with the monitor. For a failure to be declared this difference needs to be greater than a threshold, and to persist for a given period. If a failure occurs, the computer is isolated from external systems.

Even though the architecture includes four control and monitoring computers, only one of these is required to control the aircraft. Three such computers would be sufficient to meet the safety requirements (derived from system safety assessments); the fourth is included for operational reasons, as it allows a take-off to tolerate a single failed computer.

In addition to the digital electrical flight control system, there is also a mechanical back up. This is connected to the trimmable horizontal stabilizer (giving control of pitch) and the rudder (giving control of yaw). Note, however, that the safety objectives of the fly-by-wire system are defined without taking credit for the mechanical backup.

As of December 1992, the A320 had accrued 1.5 million flight hours. These included one case where, due to the combination of an air conditioning failure and a batch of ELACs being fitted with a component that did not fully meet the required temperature operating range, both ELACs were lost. Failure detection and reconfiguration was successful, with control being handled by the SECs.

5.2 Boeing 777

The following description is based on Yeh (2001).

The Primary Flight Computers (PFCs) in the Boeing 777 form a triple-triple redundant system. Safety requirements apply to PFC failures that could preclude continued safe flight and landing. The associated numerical probabilities are 10^{-10} per flight hour for both functional reliability and functional availability.

The 777 has three channels, each of which has an identical PFC. Within each channel there are three dissimilar computing lanes. This dissimilarity is based on the use of different hardware, specifically: an Intel 80486; a Motorola 68040; and an AMD 29050. This dissimilar hardware leads to a requirement for different Ada compilers.

Although the software in each computing lane was developed by different teams, all of them used the same language, namely Ada. In addition, as noted in a separate article by Yeh, Boeing's experience during the 7J7 development (which led to the 777 flight control system) was that the separate development teams had to ask so many questions to clarify the requirements that their independence was irreparably compromised (Yeh 1988).

A median value select algorithm is used on the output from the PFCs. This is supplemented by cross-lane monitoring within each PFC. There is also cross-channel consolidation between the PFCs (to maintain convergence) and cross-channel equalisation (to keep PFC channel tracking within statistically analysable bounds).

The 777 implementation is best viewed as being "replicated n-version programming". The "replicated" reflects the use of three identical PFCs. The "n-version programming" reflects that three computing lanes are used in each PFC; note, however, that the majority of dissimilarity between these lanes comes from the hardware (and associated compiler) rather than the software's design and implementation.

5.3 Summary

These examples provide valuable experience from real-world attempts to implement software design diversity. Neither of these involved any specific claim, quantified or otherwise, or the efficacy of this approach. In one case, it was simply an additional precaution; in the other only limited design diversity was actually achieved. Overall, this experience suggests design diversity cannot be relied upon as a protection measure against software CMF.

6 Evaluation Criteria

6.1 Structure

The preceding subsections have shown that, whilst it is an important issue, there is no preferred way of protecting against software CMF. Likewise, there is no commonly-agreed way of establishing the level of protection offered by a particular combination of approaches.

To overcome this issue, a set of criteria have been developed that, in combination, allow a qualitative assessment to be conducted. The selected criteria can be organised into three categories: the first addresses fault *prevention*; the second covers fault tolerance approaches against different types of potential failure-causing *inputs*; the third considers fault tolerance approaches based on the nature of the software's *outputs*.

6.2 Prevention

A single criterion relates to fault *prevention*. It considers the *software development processes* that have been used. Ideally, these would follow a widely-recognised industry standard (e.g. RTCA/DO-178C (RTCA 2011)).

6.3 Inputs

The following paragraphs summarise *input*-related criteria.

A failure related to an *unexpected external influence* from another piece of software. Such a situation may arise when one piece of (co-hosted) software runs for longer than expected, thus reducing the amount of processing time available for other software functions. Another way this situation could arise would be one application unintentionally corrupting the memory of another application. The standard way of protecting against such failures is an appropriately robust partitioning scheme.

A failure related to *a specific input bit pattern (or patterns)*. Relevant bit patterns include those corresponding to Not a Number (NaN) or Infinity (INF) (e.g. following division by zero), as well as (for example) patterns that evaluate to very small floating-point numbers. Data diversity offers a degree of protection against this potential failure mechanism. Checking inputs before they are used (i.e. defensive programming) offers a lesser, but still useful, degree of protection.

A failure related to an unlikely, *transitory environment or aircraft attitude*. Examples of such environments include crossing the Equator, flying over the North Pole, crossing the International Date Line and breaking Mach 1.0. Since the output from different (non-failing) sensors would be expected to be correlated, data diversity would not be expected to offer protection against this mechanism. Design diversity may offer some protection and, furthermore, the transitory nature of the inputs means recovery via restart (e.g. enabled via a watchdog) may also provide some protection.

A failure related to an *unlikely, but sustained environment or aircraft attitude*. Potential examples could include the aircraft having a negative forward velocity or the aircraft being inverted. As with the previous criterion, data diversity would not be expected to offer protection. In addition, the sustained nature of this criterion means that recovery via restarts would also offer little protection. Conversely, data diversity and design diversity may offer some protection. A testing regime that investigated a wide range of such situations would also help.

A failure related to *a particular sequence of events* (i.e. a failure that is caused by a specific internal state rather than a particular input). A potential example here could be a sequence of inputs that led to a navigation filter becoming unstable. The importance of the sequence means that both data diversity and design diversity would offer some protection. It also means that such issues are harder to discover via testing.

6.4 Outputs

The following paragraphs summarise *output*-related criteria.

A case where *no output* is provided. This could occur, for example, if a processor gets stuck in an infinite loop, or if the Operating System (OS) crashes. Hardware watch dogs are a standard protection approach for this criterion, because they readily facilitate failure detection and failure recovery.

A case where *a known incorrect output is provided from one computational branch (of several)*. This situation could arise, for example, due to a failing sensor giving an invalid input to a single computational branch, with the invalid nature of the input being recognised and the branch setting an output error flag. Note that the main reason why different computational branches provide different types of output is because of the presence of some diversity (either in data or design): this diversity provides some protection against this criterion. An adjudicator, or voter, that discounts the incorrect output is also required.

A case where *a known incorrect output is provided from all computational branches*. This situation could arise in a situation where a "bad" input is provided and there is no diversity

in the implementation. Being able to provide pre-defined "safe" outputs (as part of input checking) offers some protection against this situation. Restarting all the computational branches is another viable protection method (provided the aircraft can sustain controlled flight whilst the restart occurs).

A case where *an unknown incorrect output is provided from one computational branch (of several)*. Note that, as indicated above, this situation can only occur if there is some diversity in the implementation. An undetected failed sensor, which only feeds a single computational branch, could cause this situation. An adjudicator, or voter, that compares the outputs from the different branches offers some protection in this case.

A case where *an unknown incorrect output is provided from all computational branches*. This situation could arise if the platform (and hence, by extension, the embedded software) is used outside its design envelope. Diversity, and particularly design diversity, may offer some limited protection. A suitably wide, and imaginative, testing regime could also help: this is likely to benefit from a combination of software engineers and domain experts.

7 Suggested Protection Approaches

The preceding subsections have established that software CMF is an important issue, having been the cause of previous incidents and accidents, the subject of academic research, a topic in relevant standards documents and a consideration in civil aviation platforms. Nevertheless, there is no preferred way of protecting against software CMF.

It is suggested that subsystems where software CMF may pose a hazard adopt several protective approaches. These are likely to include multiple examples drawn from the following list; where appropriate additional approaches (not listed here) may also be used.

Fault Prevention, which aims to minimise the number of faults present in the software. This is generally achieved through the use of processes that apply an appropriate level of control to software development activities. RTCA/DO-178C (RTCA 2011) is an example of such processes.

Partitioning, which isolates failures in the region where they occur, rather than allowing them to cascade across the entire system. This is related to, but also supplemental to, partitioning that protects one application's memory area from the actions of another application.

Data Diversity, which involves providing different copies of software (which may or may not be identical) with different inputs. This is easiest to adopt when data diversity is naturally generated, e.g. from different sensors. Artificial data diversity can be introduced, but it can be difficult to determine the appropriate level of diversity to be used.

Input Checking, which (not surprisingly) involves checking inputs before they are used. This is a natural part of defensive programming and supports partitioning, e.g. by preventing bad data polluting subsequent computations.

Design Diversity, which typically involves using software developed by separate programming teams to perform the same function.

Failure Detection, which is concerned with identifying software failures so that they can be adequately managed and, ultimately, recovered from.

Failure Recovery, which typically involves returning the software to a "known safe" state and resuming processing.

Extensive Testing, which might involve imaginative test construction on the part of test engineers and domain experts, combined with automated test approaches, e.g. fuzzing.

For ease of reference, Table 1 indicates the relationship between evaluation criteria and the approaches suggested above. Within this table, a "++" symbol indicates a case where the approach may be expected to provide significant benefit; a "+" symbol indicates a case where moderate benefit may be expected.

Table 1 ~ Indicative Relationship between Evaluation Criteria and Suggested Approaches for Providing Protection Against Software CMF

	Fault Prevention	Partitioning	Data Diversity	Input Checking	Design Diversity	Failure Detection	Failure Recovery	Extensive Testing
Prevention: Software development processes	++							
Inputs: Unexpected external influences		++						
Inputs: Specific input bit pattern (or patterns)			++	+				+
Inputs: Unlikely, transitory environment or aircraft attitude					+	++	++	
Inputs: Unlikely, but sustained environment or aircraft attitude			+		+			++
Inputs: Particular sequence of events			++		++			
Outputs: No output						++	++	
Outputs: Known incorrect output from one computational branch			++		++	+	+	
Outputs: Known incorrect output from all computational branches				+		++	++	
Outputs: Unknown incorrect output from one computational branch			++		++	+	+	+
Outputs: Unknown incorrect output from all computational branches			+		+			++

It should be noted that Table 1 is intended to illustrate one way to substantiate an argument that protection against software CMF had been appropriately considered. In particular, the table's contents are subjective in nature and not intended to provide strict direction, nor form the basis of any algorithm to support the combination of identified techniques.

Consequently, it is not appropriate to dissect Table 1 in forensic detail. Nevertheless, there is merit in highlighting some key points.

When considering the criteria (i.e. looking across each of the rows), it is apparent that each criterion has at least one "++", with most criteria having at least one additional "+". There are two exceptions, which are criteria that only have a single "++", specifically, "Prevention: Software development processes" and "Inputs: Unexpected external influences". This is a consequence of the focus adopted for this paper. It would be possible to add significantly more detail on both of these aspects (e.g. use of formal static analysis tools, analysis of potential interference paths in multi-core processors) with a corresponding increase in symbols in the table. In addition, it is expected that these activities would be undertaken as part of any safety-related software development. Consequently, the two single "++" criteria are not judged to represent significant limitations in the approaches that are available to protect against software CMF.

When considering the approaches (i.e. looking down the columns), it is apparent that most approaches have multiple entries, both "++" and "+". There are three exceptions to this. Two of these exceptions are "Fault Prevention" and "Partitioning", which directly relate to the two criteria discussed above. The other exception is "Input Checking". As with the previous items, the apparently limited value of input checking is a direct consequence of this paper's intentionally narrow focus on software CMF. In particular, input checking is of significant wider value in safety-related software implementations.

Overall, constructions like Table 1 should help systems engineers understand, in a qualitative sense, levels of protection against software CMF. It also highlights that protection will likely have to be built from a number of approaches, and that common safety-related software approaches, e.g. fault prevention, partitioning, and input checking, are by themselves unlikely to be sufficient.

8 Closing Remarks

Software CMF is a significant issue, having been the cause of a number of accidents and incidents. It (and protection against it) has been the subject of significant academic research. Its importance is highlighted in standards and it has influenced development approaches on key civil air programmes. However, there is currently no preferred way of protecting against software CMF, nor is there an established mechanism for assessing any protection that has been implemented. To address this limitation, a set of evaluation criteria have been developed (and used on a previous platform-specific piece of work). These allow suggested protection approaches to be advanced for future air domain programmes (and other safety-related systems).

Disclaimers

This article is an overview of UK MOD sponsored research and is released for informational purposes only. The contents of this document should not be interpreted as representing the views of the UK MOD, nor should it be assumed that they reflect any current or future UK MOD policy. The information contained in this document cannot supersede any statutory or contractual requirements or liabilities and is offered without prejudice or commitment.

Acknowledgments

The authors gratefully acknowledge the comments made by the anonymous reviewers. These significantly improved the article.

References

Ammann P. E. and Knight J. C. (1988). *Data Diversity: An Approach to Software Fault Tolerance*. IEEE Trans. Computers, 37 4 (1988) 418-425.

Anderson T., Barrett P. A., Halliwell D. N. and Moulding M. R. (1985). *Software Fault Tolerance: An Evaluation*. IEEE Trans. Soft. Eng., SE-11 12 (1985) 1502-1510.

Brière D. and Traverse P. (1993). *AIRBUS A320/A330/A340 Electrical Flight Controls - A Family of Fault-Tolerant Systems*. In 23rd IEEE Int. Symp. on Fault-Tolerant Computing (FTCS-23), pages 616–623, Toulouse, France, 1993.

Hecht H. (1993). *Rare Conditions — An Important Cause of Failures*. In: Practical Paths to Assurance, Proceedings of the Eighth Annual Conference on Computer Assurance. (COMPASS '93) pp.81-85, 1993. IEEE.

IEC 61508-3:2010. *Functional Safety of Electrical/electronic/programmable electronic Safety-related Systems — Part 3: Software requirements*. IEC 61508-3, Edition 2. International Electrotechnical Commission. Geneva. 2010.

JSSSC. (2010). *Software System Safety Handbook: A Technical and Managerial Team Approach, Sub-section F.4, "Flight Controls Fail at Supersonic Transition"*. Joint Software System Safety Committee, US Department of Defense. Available from: https://ac.cto.mil/wp-content/uploads/2019/06/Joint-SW-Systems-Safety-Engineering-Handbook.pdf. Accessed 19th July 2023.

Knight J. C. and Leveson N. G. (1996). *An Experimental Evaluation of the Assumption of Independence in Multi-version Programming*. IEEE Trans. Soft. Eng., SE-12, 1 (1996) 96-109.

Laprie J-C., Arlat J., Béounes C. and Kanoun K. (1990). *Definition and Analysis of Hardware- and Software-Fault-Tolerant Architectures*. Computer, 23 7 (1990) 39-51. IEEE.

Lions J-L. (1996). *ARIANE 5: Flight 501 Failure*. Report by the Inquiry Board chaired by Prof. J-L. Lions. Available from: http://www.di.unito.it/~damiani/ariane5rep.html. Accessed 19th July 2023.

Namjoo M. and McCluskey E. J. (1995). *Watchdog Processors and Capability Checking*. Twenty-Fifth Int. Symp. on Fault-Tolerant Computing — Highlights from Twenty-Five Years, June 1995, 94-97.

Olson J. S. and Olson G. M. (2003). *Culture Surprises in Remote Software Development Teams: When in Rome doesn't help when your team crosses time zones, and your deadline doesn't*. Queue, 1(9), pp.52-59.

Powell D., Arlat J., Deswarte Y. and Kanoun K. (2011). *Tolerance of Design Faults*. In: Jones C. B. and Lloyd J. L. (editors). *Dependable and Historic Computing*. Lecture Notes in Computer Science, Volume 6875. Springer Nature, London.

Randell B. (1975). *System Structure for Software Fault Tolerance*. IEEE Trans. Soft. Eng., SE-1, 2 (1975) 220-232.

RTCA. (2011). *Software Considerations in Airborne Systems and Equipment Certification*. RTCA/DO-178C. RTCA Inc. Also available as EUROCAE Document ED-12C.

SAE. (1996). *Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment*. ARP4761, December 1996. SAE International.

Siewiorek D. P., Chillarege R. and Kalbarczyk Z. T. (2004). *Reflections on Industry Trends and Experimental Research in Dependability*. IEEE Trans. Secure and Dependable Computing, 1 2 (2004) 109-127.

TF SCS. (2022). *Licensing of Safety Critical Software for Nuclear Reactors: Common position of International Nuclear Regulators and Authorised Technical Support Organisations*. Regulator Task Force on Safety Critical Software (TF SCS). Available from: https://www.onr.org.uk/software.pdf. Accessed 20th July 2023.

Tomek L. A., Muppala J. K. and Trivedi K. S. (1993). *Modeling Correlation in Software Recovery Blocks*. IEEE Trans. Soft. Eng, 19 11 (1993) 1071-1086.

TTSB. (2021). *CI202 Occurrence Investigation*. Executive Summary of the Final Report translated into English, September 2021. Taiwan Transport Safety Board. Available from: https://www.ttsb.gov.tw/media/4913/ci202_executive-summary_release.pdf. Accessed 19th July 2023.

UKMOD. (2020). *Certification Specifications for Airworthiness: Part 1 — Fixed Wing Combat Air Systems*. DEF.STAN.00-970, Issue 17, August 2020.

USDoD. (2014). *Airworthiness Certification Criteria.* MIL-HDBK-516C, December 2014.

USNRC. (2011). *Criteria for Use of Computers in Safety Systems of Nuclear Power Plants*. Regulatory Guide 1.152, Revision 3, July 2011. U.S. Nuclear Regulatory Commission.

van der Meulen M. J. P. and Revilla M. A. (2006). *The Effectiveness of Software Diversity in a Large Population of Programs*. IEEE Trans. Soft. Eng., 34, 6 (2006) 753-764.

Voas J., Ghosh A., Charron F. and Kassab L. (1997). *Reducing Uncertainty About Common-Mode Failures*. In: Proceedings of the Eighth International Symposium on Software Reliability Engineering, 1997, pp. 308–319.

Walmsley R., Anderson T., Brendish C., McDermid J. A., Rolfe M., Sultana J., Swan M. and Toms M. (2015). *NATS System Failure 12 December 2014 — Final Report*. Independent Enquiry Panel chaired by Sir Robert Walmsley KCB. Available from: https://www.nats.aero/wp-content/uploads/2015/05/Independent-Enquiry-Final-Report-2.0.pdf. Accessed 19th July 2023.

Yeh Y. C. (1988). *Design Considerations in Boeing 777 Fly-By-Wire Computers*. In 3rd IEEE Int. Symp. on High-Assurance Systems Engineering, pages 64-72, 1988.

Yeh Y.C. (2001). *Safety Critical Avionics for the 777 Primary Flight Controls System*. Digital Avionics Systems Conference, 2001 (20th DASC).

ISSN 2754-1118 (Online)
ISSN 2753-6599 (Print)

An IEC 61508 Viewpoint on the Safety Assessment of Railway Control Systems

Derek Fowler[1] and Alasdair Graebner[2]

1. Independent Safety Engineering Consultant, Reading, UK
2. Railway System Engineer, UK

Abstract

An article entitled "An IEC 61508 Viewpoint on System Safety in the Transport Sector" in the July 2022 edition of the Safety-critical Systems eJournal, proposed a way of thinking about the safety assessment of transportation systems that is based on the fundamental principles of international functional-safety standard IEC 61508. Then, in a second article (January 2023), an operational example from Air Traffic Management (ATM) was used to outline how an IEC 61508 approach to safety assessment could be applied to the ATM sector. Now, in this article, the example of a new, moving-block Automatic Train Control system, for a hypothetical Metro, is used to outline how an IEC 61508 approach to safety assessment could be applied to the safety assessment of railway control systems in general.

1 Introduction

1.1 Background

IEC 61508 — "*Functional Safety of Electrical/electronic/programmable electronic Safety-related Systems*" (IEC 2010) — is probably the most widely-accepted, international generic standard on functional safety. Although its ancestry can be traced back to process industries, the intention behind the Standard has always been to provide a solid, comprehensive basis for adaptation, as necessary, to meet the needs of a wide range of industry sectors.

In the first of three related articles, Fowler (2022) proposed 'a way of thinking' about the assessment of the various safety-related systems deployed in the Transport sector — especially commercial-aviation and rail applications — based on the key principles and safety lifecycle set out in Parts 1 and 4 of IEC 61508.

A follow-up article (Fowler and Fota, 2023) then took the example of an innovative Air Traffic Management (ATM) operational concept and used it to outline how an IEC 61508 approach to safety assessment could be applied effectively to the ATM sector, and what the results thereof might look like, starting with the concept of the traffic in the airspace being (what IEC 61508 calls) the Equipment Under Control (EUC).

This article now takes a similar approach but with the example of a new, moving-block Automatic Train Control (ATC) system for a hypothetical Metro; this is used to outline how an IEC 61508 approach to safety assessment could be applied effectively to the

© Derek Fowler and Alasdair Graebner 2023.

Railway sector in general, and what the results thereof might look like, starting with the concept of the movement of trains around the railway being the EUC.

It is important, at this stage, to clarify five features of the article, as follows:

Firstly, it is generic in its approach and is not a case study of a particular project.

Secondly, due to the vast amount of detail that would be involved in a full system safety assessment of this kind, it is not an exhaustive, rigorous study — rather, only examples of the outputs of the relevant phases of the IEC 61508 safety lifecycle are provided, sufficient to give insights into the processes involved.

Thirdly, it is not the intention to prescribe IEC 61508-compliant processes for railway applications — rather, it is to use the lifecycle cycle model from Part 1 of IEC 61508 (IEC 61508-1) to shape thinking about system safety assessments away from a mindset that, in the past at least, "*focussed too much on system reliability and not enough on system functionality, contrary to, inter alia, the most basic principles of the international functional-safety standard IEC 61508*" (Fowler 2022).

Fourthly, in line with the approach taken in the two previous articles in this series, i.e. Fowler (2022) and Fowler & Fota (2023), it is not the intention to carry out a detailed compliance assessment of current railway safety standards and practices against IEC 61508. That said, it is worth drawing attention here to the following, rather bold, statement in the latest version of EN 50126-1 (CENELEC 2017):

> *"EN 50126 forms part of the railway sector specific application of IEC 61508. Meeting the requirements in this European Standard, together with the requirements of other suitable standards, is sufficient to ensure that additional compliance to IEC 61508 does not need to be demonstrated",*

and to the findings of Fowler (2015), which suggested that the version of EN 50126, applicable at that time (1999), fell well short of compliance with the most basic principle of IEC 61508, which is outlined in Sub-section 1.2 below. Suffice it to say that, to date, we have found no evidence to show that the CENELEC (2017) statement is justified in today's, increasingly-automated railway environment, but would be happy to hear from any readers as to how, and where, current railway standards do meet the particular IEC 61508-1 requirements presented in this paper, should that be the case.

Fifthly, whereas numerous references are made to parts of IEC 62260, *Urban Guided Transport Management and Command/Control Systems (UGTMS)*, the use of this standard (IEC 2014) is simply and solely as a convenient source of information that would otherwise have to be derived from scratch, e.g. the very large number of generic system functions described in Sub-section 4.6 below.

1.2 The IEC 61508 Viewpoint

Part 4 of IEC 61508 (IEC 61508-4) defines Functional Safety as being:

> *"that part of the overall safety relating to the EUC / EUC Control System that depends on the correct functioning of the safety-related systems and other risk-reduction measures".*

It is founded on the IEC 61508 fundamental principle that:

- where there exists an Equipment Under Control (EUC)[1], with its associated Control System[2], which is *inherently* hazardous to the environment in which it operates; then
- Safety-Related Systems (SRSs)[3] and/or Other Risk-reduction Measures (ORRMs) need to be developed, in order to *reduce*, to a tolerable level, the inherent risk presented by the EUC.

This simple principle of risk reduction is fundamental to IEC 61508 and is illustrated in the risk graph of Figure 1, which itself is derived directly from Figure A.1 of Part 5 of IEC 61508 (Fowler 2022).

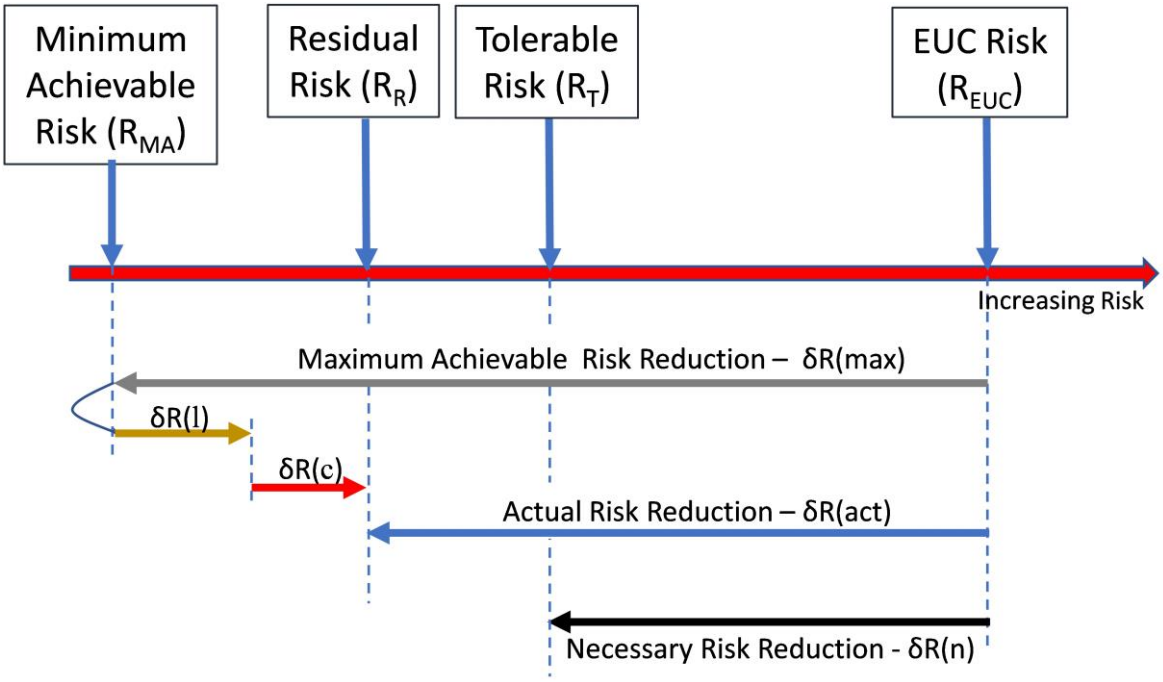

Figure 1 ~ Risk Graph

The inherent (or unmitigated) EUC Risk (R_{EUC}), provides a (usually theoretical) reference point that takes *no* account of the possible risk reduction afforded by any SRSs / ORRMs, and Necessary Risk Reduction (δR(c)) is the amount of risk reduction that must be *achieved* by the SRSs / ORRMs in order to ensure that the Tolerable Risk is not exceeded.

Residual Risk (R_R) is the risk that is *actually* achieved for the EUC, with *full* account of the risk reduction afforded by the SRSs / ORRMs now taken into account, and depends on three properties of those SRSs / ORRMs:

- their functionality and performance, which determine the maximum achievable risk reduction δR(max), if (theoretically) the SRSs / ORRMs never failed;
- their reliability, in terms of the likelihood of their failure to function at all, and thereby reducing the achievable risk reduction by an amount δR(l); and

[1] "*Equipment, machinery, apparatus or plant used for manufacturing, process, transportation, medical or other activities*" (IEC 61508-4)

[2] "*System that responds to input signals from the [EUC] … and/or from an operator, and generates output signals causing the EUC to operate in the desired manner* " (IEC 61508-4) but without specific regard to the safety of that operation.

[3] Designated system that both: implements the required safety functions necessary to achieve or maintain a safe state for the EUC; and is intended to achieve, on its own or with other safety-related systems and 'other risk-reduction measures', the necessary safety integrity for the required safety functions (IEC 61508-4).

- their integrity, in terms of the likelihood of their operating corruptly (i.e. spuriously or incorrectly), and thereby introducing a new source of risk of δR(c).

Hence, given that the unmitigated EUC Risk (R_{EUC}) would be *many* orders of magnitude greater than the Tolerable Risk (R_T), the inescapable conclusion is that we must first specify SRSs / ORRMs that are proven to have sufficient functionality and performance to achieve a risk *reduction* in the absence of failure (δR(max)) that is greater than what is "necessary" (δR(n)), before considering any risk i*ncrease* caused by such failures.

That is the IEC 61508 viewpoint, on which the remainder of this article is based.

2 Scope

Like Fowler (2022), the scope of this article is limited to the initial phases of the IEC 61508 safety lifecycle, which result in the specification of detailed *functional safety requirements*[4] and *safety integrity requirements* necessary and sufficient for the subject SRSs / ORRMs to achieve a tolerable level of risk for the EUC.

The relevant phases are shown in Figure 2, overleaf, which is based on Figure 2 of IEC 61508-1 (IEC 2010), with the following modifications:

- Phases 10 and 12 to 14 have been omitted since they address requirements realisation and, therefore, fall outside the scope of this article;
- Phases 6 to 8 have been omitted as they cover only planning for the realisation phases;
- a summary of the main outputs of each phase has been added; and
- the specification of safety requirements for ORRMs, in Phase 11, falls *within* the scope of this article, even though it is outside the scope of IEC 61508 itself.

It should be noted also that IEC 61508's use of the term "*E/E/PE (System) — Electrical/Electronic/Programmable Electronic (System)*" — was felt to be too specific and limiting for the purposes of this article; therefore, the more general term "safety-related system (SRS)" is used instead herein so as to allow human and procedural elements to be included as well as (and possibly instead of) technical equipment[5].

In line with Fowler (2022), it is expected that the resulting system safety requirements would be sufficient to ensure that:

- under all *normal* operating conditions[6], a fully-functioning railway control system would be capable of mitigating all potential EUC hazards[7], such that at least a tolerable level of EUC risk would be achieved;
- a fully-functioning railway control system would be able to continue to mitigate potential EUC hazards, under all *abnormal* operating conditions[8] without a significant increase in the achievable level of EUC risk;

[4] The term *functional safety requirements* was coined in Fowler (2022) in preference to the (arguably ambiguous) IEC 61508 term of *safety functions requirements*; it covers safety requirements for both functionality (what has to be done) and performance (how well it has to been done).

[5] IEC 61508-1, Sub-section 1.2, Note 2 states that "…a person can form part of a safety-related system".

[6] That is all those conditions that are expected to occur on a day-to-day basis (Fowler 2022).

[7] That is, those hazards that are *inherent* in railway operations *before* any safety-related systems are provided in order to mitigate them (Fowler 2022)

[8] That is those conditions that are expected to occur less frequently but under which the ATC system is expected to continue without significant degradation of its primary functionality or performance.

- the causes and consequences of *failure* conditions, within the subject railway control system, are controlled such that the overall achievable level of EUC risk would remain at least tolerable.

As we work herein through these lifecycle phases for the subject railway operations, it might appear that some of the steps could be simplified by, for example, subsuming them into other steps. Indeed, IEC 61508 allows for this to be done, where applicable, but, for the purposes of this article, we decided to adhere exactly to the lifecycle, which was detailed previously in Fowler (2022), except where indicated otherwise in the Sections below.

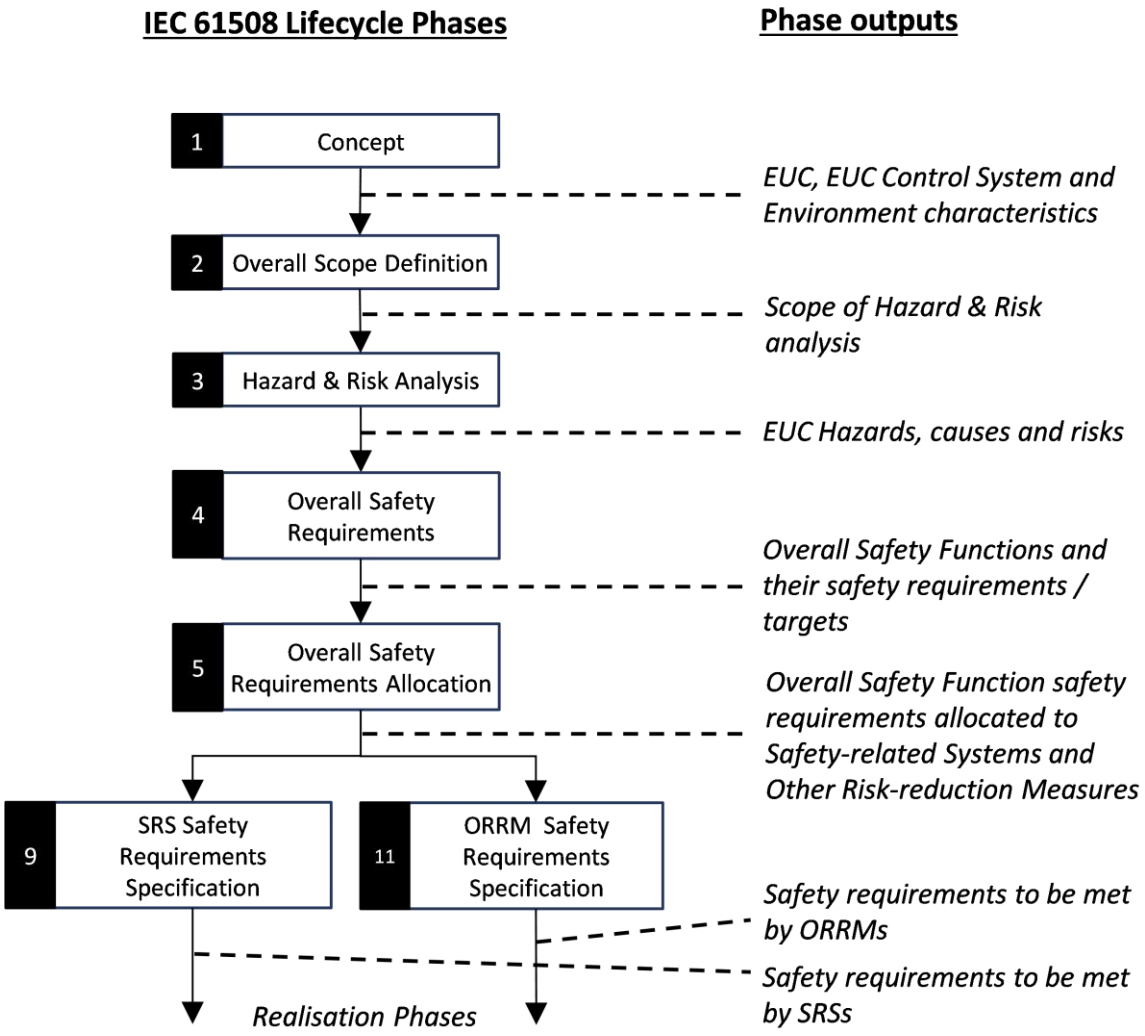

Figure 2 ~ Applicable IEC 61508 Overall Safety Lifecycle

3 Operational Context

It is assumed that a new Automatic Train Control (ATC) system will replace a conventional, fixed-block signalling system (including automatic train protection), on a hypothetical existing Metro.

The ATC system will be based on the Communications-based Train Control (RailSystem 2022) concept in which real-time, train-control information (based on radio communications), and moving-block signalling principles, are used to increase line capacity (by reducing the headway between trains travelling on the same line), and to minimise the amount of trackside equipment, *without* any degradation in the safety of the railway operations.

Conceptually, railway control systems (including our ATC system) can be thought of as comprising the following:

- Automatic Train Supervision (ATS): ensures the safety of all trains by continuously detecting the presence, or absence, of trains and (where applicable) transmitting safety speed and distance data from the wayside; by applying the correct settings to infrastructure and signalling assets; and by directing both the movement and movement authority for each train formed to deliver a timetabled service as requested by ATR;
- Automatic Train Regulation (ATR): interprets the timetable and delivers the planned service; continually monitors the progress of trains, detecting when trains are running "off timetable"; and regulates the progress of a train, or trains, to bring services back in line with the timetable;
- Automatic Train Operation (ATO): receives information from the signalling system regarding movement authority and required speed profile; and causes the train to proceed when in an automatic driving mode; and
- Automatic Train Protection (ATP): continuously compares the actual train speed with the safety speed limit applicable at that time for the section occupied by the train; and causes the train to emergency brake in the event of an infringement.

In general, ATP and ATS are considered to be "vital" systems on the basis that their primary purpose is *accident prevention* and on which the safety of the railway critically depends — in IEC 61508 terms, they would fall into the category of safety-related systems. The purpose of ATO and ATR, on the other hand, is *primarily* the efficient running of the railway though, as we will see in Sub-section 4.6.6 below, ATR and/or ATO might also make some positive contribution to safety.

The level of automation for a fully-functioning, passenger-carrying train is assumed to be *semi-automated,* or GOA3, which is defined in the UGTMS standard (IEC 2014a) as follows:

> *"The driver is in the front cabin of the train observing the guideway*[9] *and stops the train in the case of a hazardous situation. Acceleration and braking are automated, and the speed is supervised continuously by the* [ATC] *system. Safe departure of the train from the station is the responsibility of the operations staff (door opening and closing may be done automatically."*

In effect, the control of train movements, i.e. our EUC, is fully automated whereas other areas of safety concern, such as managing the platform-train interface, are not. Even so, the choice of a GOA3 system here still presents a major challenge for this, and all similar, advanced-technology safety assessments[10]; as pointed out in Fowler (2022), regarding the introduction of "*self-driving*" cars, it would be naïve to assume that replacing ("unreliable") human operators by supposedly more-reliable computer-based systems

[9] Also known as the permanent way

[10] Clearly, removing the driver from the cab (GOA4) would provide an even greater challenge; it would, however, also introduce further complexity that would be difficult to handle within this article, without adding much to its key message.

would lead directly to fewer accidents without *first* assessing whether those systems would be capable of matching the traditional skills and experience of humans in equivalent transportation roles.

4 Safety Assessment

4.1 Concept (IEC 61508-1 Phase 1)

4.1.1 Aim

The aim of Phase 1 is to gather as much information about the *Equipment Under Control* (EUC), its *Environment*, and the *EUC Control System*, as is necessary and sufficient to enable the other safety lifecycle activities to be satisfactorily carried out.

It is important to note that, as an enabling activity, this would be a precursor to, but not form part of, the safety assessment *per se* and would require substantial operational and systems-engineering specialist input, relevant to each specific application. In practice, such material may be found in a typical *Operational Concept* document.

4.1.2 EUC

As with any other railway-signalling application, we can understand the EUC as being, in general, the movement of trains around the rail network, for whatever purpose. This understanding is consistent with the core IEC 61508 principle that the EUC is the main source of hazards, the mitigation of which Safety Related Systems (SRSs) and/or Other Risk-reduction Measures (ORRMs) are provided, to achieve a tolerable level of EUC risk.

The key inherent properties of the EUC that we would need for a full safety assessment are as follows:

- Train types:
  - ATC-equipped passenger-carrying trains
  - ATC-equipped engineering trains, in various formations
  - “Alien” trains, i.e. not ATC equipped;
- Passenger-carrying train properties:
  - Configuration: e.g. 7 cars, with open gangways, operating as a single unit
  - Train length
  - Tare mass
  - Power: e.g. electric 3rd rail 750V DC, running rail return
  - In service motoring and braking characteristics
  - Emergency braking capability
  - Passenger capacity;
- Railway system properties:
  - Fleet size
  - Target peak trains per hour (timetabled in each direction).

4.1.3 Environment

IEC 61508 defines the "environment" for the EUC in terms that include its physical, operating, legal and maintenance properties.

The environment properties for the subject ATC operations, which determine the functionality, performance and integrity *required* of the ATC system, usually include:

- Weather conditions, e.g. visibility and rail icing, and frequencies thereof;
- Poor rail adhesion, e.g. leaf fall on to track;
- Flood risk, which for the purposes of this article is assumed to be negligible;
- Track parameters, as follows:
    - total running length (excluding depots & sidings)
    - percentage of the track that is underground
    - maximum line (design) speed
    - number and details of stations, above and below the surface
    - number and details of depots
    - number and details of sidings
    - types and layout of demandable elements, e.g. points & controlled crossings, and non-demandable elements, e.g. diamond crossings
    - availability of secondary train detection and wayside signals for alien trains;
- Properties of individual stations, including:
    - platform lengths
    - the presence (or otherwise) of platform-screen and platform-end doors (both are assumed to be present)
    - the platform-train interface (PTI).

4.1.4 EUC Control System

Given the above interpretation of the EUC as being the movement of trains around the rail network, we can view the EUC Control System as being the functional system (comprising people, procedures and equipment), whose primary aim is to control that movement in the desired manner *and* facilitate the embarkation and disembarkation of passengers.

In its basic form, IEC 61508 generally makes a distinction between the EUC Control System and the Safety Related Systems (SRSs) that are required additionally in order to reduce the risks that are inherent in the operation of the EUC.

Fortunately, IEC 61508-1 also permits parts, or all, of an EUC Control System to be considered to be safety-related, *provided* they are subject to the appropriate requirements of the Standard (Fowler 2022) and, therefore, the need for rigid distinctions to be drawn between what is vital and non-vital is obviated.

What *is* important from an IEC 61508 perspective, is the relationship between the EUC and the EUC Control System, which is indicated in Figure 3.

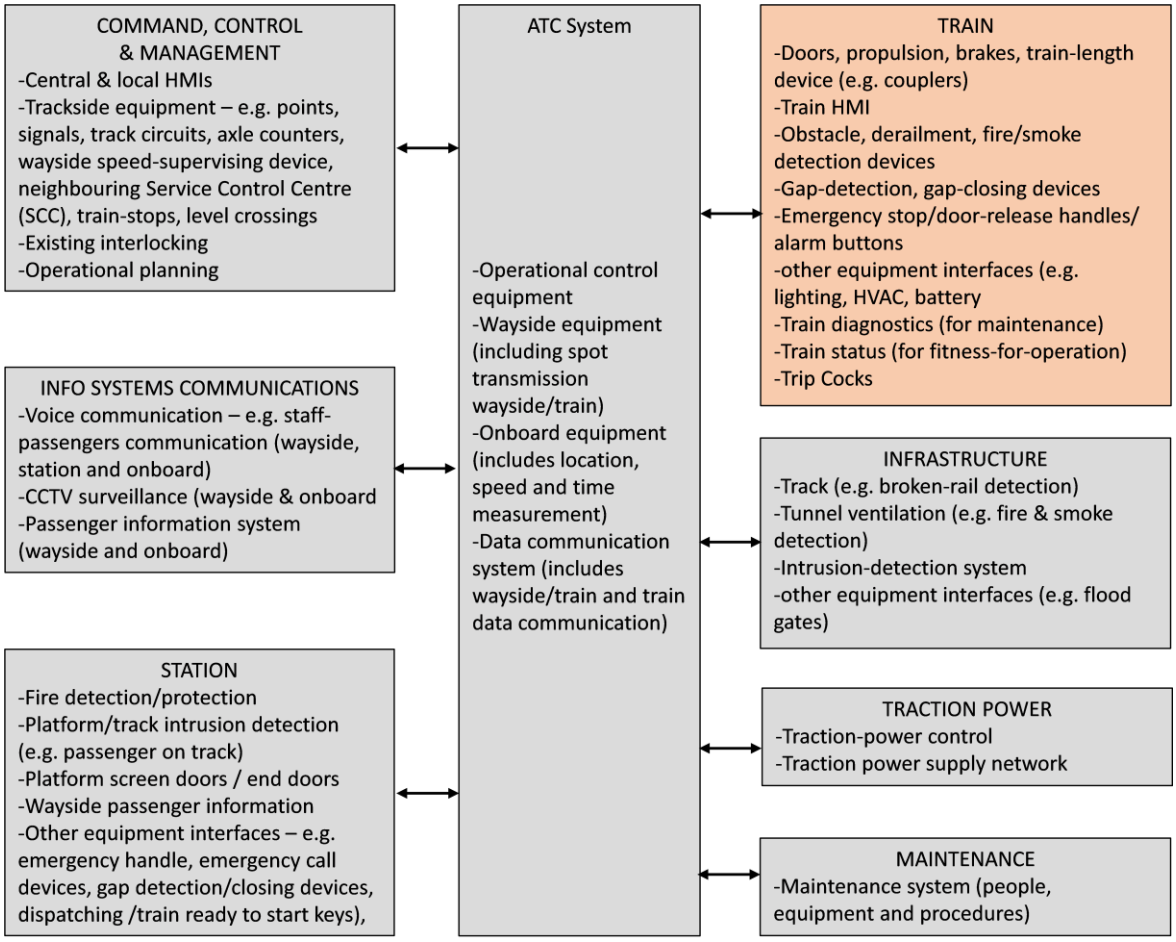

Figure 3 ~ Logical System Relationships

Strictly speaking, Figure 3 necessarily shows only the physical attributes of individual trains whereas, the EUC is defined, at the more conceptual level of Sub-section 4.1.2, as the general *movement* of such trains around the rail network. It is important to note also that, although the allocation of the detailed items in Figure 3 to the various system elements might vary slightly, case-to-case, the following provisions always apply:

- the ATC element is the *only* one for which Safety Requirements are actually derived;
- Safety Requirements are *not* derived for any items that form part of the basic Train vehicular element[11] since it is assumed that their safety would have been established through a prior safety assessment / monitoring process of the old signalling system;
- Safety Requirements are also *not* derived for any items that form part of any *non*-ATC EUC Control System elements; however, in the event that the safety of ATC operations depends on an *assumption* of the safety integrity of any such items being more stringent than 10^{-5} dangerous failure per hour (or low-demand / on-demand equivalent) then those items must be deemed to be safety-related — see Sub-section 3.5.2 of Fowler (2022).

[11] *Except* for any train-mounted ATC items

4.2 Overall Scope Definition (IEC 61508-1 Phase 2)

4.2.1 Aim and Objectives

The aim of Phase 2 is to define the scope of the Hazard and Risk Analysis, which will be carried out in Phase 3.

It seeks to achieve that aim through determining the boundary of the EUC / EUC Control System and its Operational Environment and, within those constraints, specifying the scope of the Hazard and Risk Analysis.

This is particularly important when, as we are doing in this article, assessing the safety of a change to an existing railway operation and/or systems so as to identify, and exclude, the unnecessary safety assessment of those elements that are not affected by the change. It should be noted, however, that we can do this only in general terms herein because of the necessarily generic nature of the operational context for which this example safety assessment is being carried out.

4.2.2 Boundary Constraints

For the purposes of this safety assessment of ATC operations, the train movements, which constitutes the EUC, are *only* those that:

- occur *within* the specified ATC signalling-system boundary; or
- involve the *transfer* of trains to/from any adjacent signalling areas, in accordance with the required boundary conditions; or
- involve the *transfer* of trains to/from any adjacent, non-signalled areas, e.g. depots, in accordance with the required boundary conditions.

4.2.3 Scope of the Hazard and Risk Analysis

Subject to the above boundary constraints, the scope of the Hazard and Risk Analysis shall *include* all hazardous events that are *inherent* in:

- the general movement of trains under normal, abnormal and failure conditions;
- the transfer of passengers, on and off the train, at a station, from and to the platform;
- the necessary presence of maintenance staff and equipment on the track;
- the necessary presence of passengers on the track during, for example, evacuation from a train or station; and
- interactions between trains and road users at level crossings.

The scope shall *exclude* any other hazardous events, i.e. those that do not fall within the scope of the bullet list above and/or occur outside the boundaries defined in Sub-section 4.2.2.

4.3 Hazard and Risk Analysis (IEC 61508-1 Phase 3)

4.3.1 Aim

The aim of Phase 3 is to determine, and characterise, all the hazards and risks associated with the EUC[12], in the stated Operational Environment, and within the scope already identified in Phase 2.

Note: it is acknowledged that these EUC hazards (and some of the detail that follows, up to and including Sub-section 4.4.2 below), which are not specific to ATC operations, might have already been identified and documented adequately in, say, a safety case for the current (fixed-block) railway operations. For the purposes of this article, however, we will present the analysis as if no such previous work had been done.

4.3.2 EUC Hazard Identification

The objective here is to determine the hazards relating to the EUC, within the scope defined in Sub-section 4.2.

From the IEC 61508 definition of a hazard, which can be paraphrased as "a potential **source** of harm, i.e. death, physical injury or damage to the health of people or damage to property or the environment" (Fowler 2022), it follows that we must first identify the types of harmful **outcome**, i.e. accident, that fall within the signalling area's general area of responsibility and specifically within the above scope of ATC operations.

Table 1 suggests various accident types relevant to railway operations, on the subject railway and, in each case, would involve death or serious injury to one or more of those on board a train or to the workforce or members of the public on, or in the vicinity of, the track.

Table 1 ~ Accident Types Relevant to Railway Operations

ID	Accident Type	Description
A#1	Collision between trains	All collisions between trains except where preceded by de-railment of at least one of the trains involved
A#2	Derailment of a train	Unintentional departure of a train from the track
A#3	Collision between train and road users	Train collides with road vehicle, cyclists and/or pedestrians on a level crossing
A#4	Collision between train and non-fixed obstacle(s) on the track	Train collides with non-fixed objects (including debris, members of the public or large animals) on the track
A#5	Collision between train and personnel on the track	Train collides with workforce (including their equipment or vehicles) or disembarked passengers who are on the track

[12] Strictly speaking, IEC 61508 includes "EUC Control System Hazards" here as well. We have taken the view that, for ATC, failures of the EUC Control System are among the *causes* of EUC hazards

ID	Accident Type	Description
A#6	Collision between train and fixed structure	Train collides with a fixed structure (including buffer stop), except as a result of either a derailment (A#2) or failure of such structure (A#4)
A#7	Passengers falls on board a train	Passengers injured by falling due to sudden, violent acceleration or deceleration of the train.
A#8	Passenger falls from train on to track	Passenger deaths / serious injuries due to falling from stationary or moving train, on to track
A#9	Passenger slips or trips when getting on or off a train at platform	Passenger deaths / serious injuries due to slips / trips during embarkation / disembarkation at platform, including dragging due to becoming caught in the closing train doors
A#10	Fire on board a train, in a station or trackside	Passenger deaths / serious injuries due to exposure to heat and/or smoke inhalation from a fire on a train, in a station or trackside
A#11	Fatal or serious electrical injury to passengers or workforce	Passenger or workforce deaths / serious injuries due to exposure to lethal voltages or arcing – resulting injuries include electric shock and burns from contact with live parts, or injury from exposure to arcing

The hazards derived from the above, and in relation to what are seen to be the most credible accident outcome(s), are shown in Table 2 and were adapted from the set of "core" railway hazards derived in doctoral research carried out by Ivan Lucic (Lucic 2015); all of these hazards are *inherent* in railway operations, in the stated Operational Environment, and exist *before* any form of hazard mitigation has been applied.

In the specific case of Hp#1 to Hp#12, the hazards apply directly to the EUC, i.e. the movement of passenger-carrying and engineering trains, and their mitigation places *direct* demands on the safety functionality of the ATC system.

The remaining five hazards, which have a much less direct impact on the required functionality of the ATC system, are not considered to be EUC hazards but will be addressed as part of the analysis of *abnormal* operating conditions, for which the ability of the ATC system to react appropriately will still have to be demonstrated (Sub-section 4.6.5 below).

Table 2 ~ Hazards Inherent in Railway Operations

ID	Hazard	Related Accident(s)
EUC Hazards		
Hp#1	Conflict (1) between any pair of train trajectories (2)	A#1
Hp#2	Conflict (1) between a train's trajectory (2) and track configuration	A#1, A#2, A#3
Hp#3	Train speed exceeding capabilities of the track infrastructure and/or train	A#2, A#6
Hp#4	High and/or uneven acceleration / deceleration of a train	A#7

ID	Hazard	Related Accident(s)
Hp#5	Conflict (1) between train profile and fixed structure, *except* as the result of excessive train speed (Hp#3) or damage to structure (Hp#10)	A#6
Hp#6	Conflict (1) between a train's trajectory (2) and non-fixed obstacles or unauthorised persons on track	A#4
Hp#7	Conflict (1) between a train's trajectory (2) and workforce / vehicles on track	A#5
Hp#8	Passengers attempt to exit train outside a station	A#8
Hp#9	Passenger embarkation / disembarkation at platform	A#9
Hp#10	Structural failure of track elements, tunnels, bridges etc	A#4
Hp#11	Personnel exposure to potentially lethal voltage	A#11
Hp#12	Passengers too close to, or fall/jump off, platform edge	A#4, A#11
Other Inherent Hazards		
Hp#13	Passenger evacuation outside platform	A#5, A#11
Hp#14	Train encounters adverse rail-surface conditions	A#1 to A#5
Hp#15	Conflict between a train's trajectory (2) and trackside fire	A#10
Hp#16	Station fire / other emergency on a station	A#10, A#4
Hp#17	Fire, or other emergency, on board a train	A#10

Notes:

1. For the specific meanings of "Conflict" in each case, see Appendix A
2. Conceptually, a train's "trajectory" is the path and speed profile that the train intends to follow at any point in time, and in the absence of any contrary instructions or information.

IEC 61508 requires that the sequence of events be described for each EUC hazard at this stage in the process, but to do so exhaustively would normally be impracticable for railway operations, because of the sheer number of causal factors involved. What we can usefully do, however, is to describe in general terms the precursor to each hazardous event, and this is included in the more detailed hazard descriptions at Appendix A; we then leave it to the modelling approach described in Sub-section 4.6 below, which does capture how such states are arrived at in the first place, and thus fully satisfy this IEC 61508 requirement.

Of course, what we have not said thus far is anything about the probability that each EUC hazardous event would lead to the related accident except, that the probability would, by definition, be finite. That is addressed next.

4.3.3 EUC Risks

The objective here is to determine the EUC Risks from two perspectives.

Firstly, for each accident type identified in Table 1, the *tolerable* level of EUC Risk must be identified; since the accident types would be unchanged from the previous, fixed-block

operations, it is reasonable to assume, at this stage, that what are *deemed* to be relevant tolerable levels of risk would already have been promulgated.

Secondly, for each hazardous event identified in Table 2, IEC 61508-1 suggests that the *expected* value of unmitigated EUC Risk be estimated at this stage, i.e. *without* taking into account the possible risk reduction afforded by any Safety-related Systems, or any Other Risk-reduction Measures, that would be developed subsequently for that purpose.

However, as discussed in Fowler (2022), there are significant problems in estimating such values of unmitigated EUC risk for complex applications typical of the transport sector; fortunately, as explained in Sub-section 4.4 below, the determination of absolute EUC Risk is not actually necessary in practice, *provided* the associated concept of Necessary Risk Reduction is adhered to in the determination of Overall Safety Requirements.

4.4 Overall Safety Requirements (IEC 61508-1 Phase 4)

4.4.1 Aim

The aim of Phase 4 is to produce a specification of the Overall Safety Requirements for each Overall Safety Function (OSF) in order to achieve the required level of functional safety.

The specification covers both the functional safety requirements (FSRs) and safety integrity requirements (SIRs) for the OSFs although, as we will see in Sub-section 4.4.4 below, IEC 61508's use of the term *safety integrity requirements* at this level is somewhat confusing!

4.4.2 Overall Safety Function Identification

The objective here is to identify a set of OSFs, based on the EUC hazardous events derived from the hazard and risk analysis of Phase 3.

According to IEC 61508, an overall safety function is the highest-level abstraction of the "*Means of achieving, or maintaining, a safe state for the EUC, in respect of a specific hazardous event*"[13], whereas, for the Rail sector, the relationships between accidents and hazards (as shown above) is "many-to many", as is the relationship between EUC hazards and the OSFs that are intended to mitigate them.

However, as found in Fowler and Fota (2023) for the Air Traffic Management sector, this is not an insurmountable problem, and the set of OSFs proposed in Table 3 otherwise seems to fit the above definition of an OSF very well.

Table 3 ~ Overall Safety Functions

OSF ID	OSF Title	Related EUC Hazards
OSF#1	Establish & Protect a Safe Route for each Train Movement	Hp#1, Hp#2, Hp#5
OSF#2	Apply & Maintain Safe Separation between Trains	Hp#1

[13] IEC 61508 terminology can be a bit confusing here (Fowler 2022). Hierarchically, an *Overall* Safety Function can be realised as one or more Safety Related Systems and/or one or more Other Risk-reduction Measures, and a Safety Related System can be realised as one or more Safety Functions.

OSF ID	OSF Title	Related EUC Hazards
OSF#3	Enforce Safe Speed Limits for Trains	Hp#3
OSF#4	Provide Safe Passenger Embarkation / Disembarkation	Hp#6. Hp#8, Hp#9, Hp#12
OSF#5	Provide Safe Maintenance Access to Track	Hp#7
OSF#6	Ensure that the Guideway is Safe for Train Passage	Hp#6, Hp#10
OSF#7	Ensure Safe Acceleration & Braking	Hp#4
OSF#8	Ensure Safety of Traction Power Supply	Hp#11

4.4.3 Determine the Functional Safety Requirements for each Overall Safety Function

This step involves the determination of what is required functionally from each of the above OSFs. The resulting Overall Safety Requirements (OSRs) are based on *normal* operational conditions, as described in Section 2, and cover those items that are necessary and sufficient to ensure the safety of ATC operations, in the absence of *failure* and of *abnormal* operating conditions.

The properties shown in Table 4 are what is required of the respective OSFs in order to avoid, and / or mitigate the consequences of, the EUC hazards shown in Table 2.

Table 4 ~ Overall Functional Safety Requirements for Normal Operations

Requirement ID	Requirement Description	Related EUC Hazard
OSF#1	**Establish & Protect a Safe Route for each Train Movement**	
OSR1.1	A train shall *not* be authorised to enter a route unless, *and until*, the route is set and locked in a safe condition (see OSR1.2) and reserved, for *that* train	Hp#1, Hp#2
OSR1.2	A route shall be defined by: - the *route origin* (the location for which authorisation for a train to enter the route shall be given) and the *route destination* (the location at which the movement authority ceases); - all the route elements between the route origin and route destination, which are to be traversed by the train; - route elements of overlap, which are reserved for safety reasons in case of deviations from an authorized train movement; - route elements in the flank-protection area, which prevent or detect unauthorised flank movement; - the authorised direction of travel for the train.	Hp#1, Hp#2

Requirement ID	Requirement Description	Related EUC Hazard
OSR1.3	A route shall be considered as safe if, *and only if*, - every requested element of the guideway is locked in the required position such that concurrent use by another train is avoided entirely; *and* - road vehicles (and other road users) are prevented from occupying a level crossing on the route, prior to, and during, train passage; *and* - every requested elements of the guideway that provide flank protection is locked in the required position	Hp#1 Hp#2 Hp#1
OSR1.4	A route element shall *not* be released and reset for another train until the previous train has cleared that element of the route	Hp#1
OSR1.5	It shall *not* be possible to route a train through or past a fixed structure whose gauge is incompatible with the kinetic envelope of that train	Hp#5
OSR1.6	The probability of a train overrunning its limit of safe route shall not exceed 10^{-9} per operating hour	Hp#1, Hp#2
OSR1.7	It shall not be possible to run a train beyond the end of route, or into an area controlled by another signalling system without permission	Hp#1, Hp#2
OSF#2	**Apply & Maintain Safe Separation between Trains**	
OSR2.1	A safe distance between following trains (see OSR2.2) shall be maintained at all times	Hp#1
OSR2.2	*Safe distance* shall be based upon the principle of an instantaneous stop of the preceding train *and* on the ability of the following train to be braked to a halt in time to avoid a collision	Hp#1
OSR2.3	The *safe distance* shall be sufficient to ensure that, under *normal* operating conditions, the probability of a train being unable to stop before colliding with the leading train shall not exceed 10^{-9} per operating hour	Hp#1
OSF#3	**Enforce Safe Speed Limits on Trains**	
OSR3.1	A train's actual speed shall not exceed its safe speed (see OSR3.2) at anytime	Hp#3

Requirement ID	Requirement Description	Related EUC Hazard
OSR3.2	The safe speed shall be the least of: - the speed above which it would not be possible to bring the train to a halt before reaching the limit of its Movement Authority, without the use of emergency braking; and - any permanent and temporary speed restrictions applicable to the track infrastructure within the train's movement authority; and - any temporary speed restrictions applied in response to degraded environmental conditions within the train's movement authority; and - any permanent or temporary speed restrictions applicable to the train itself.	Hp#3
OSR3.3	Permanent speed restrictions shall be determined on the basis of what would be tolerably safe for the train type, state and track-infrastructure geometry	Hp#3
OSR3.4	Temporary speed restrictions shall be determined on the basis of what would be tolerably safe under the actual conditions of the train, track or environment	Hp#3
OSR3.5	On approaching an area with a lower speed limit, a train shall have reduced its speed to the new speed limit prior to entry into that area	Hp#3
OSR3.6	All speed restrictions for the track infrastructure shall be applied to the whole length of the train	Hp#3
OSR3.7	The probability of a train exceeding its safe speed, by an amount sufficient to cause derailment, or other major accident, shall not exceed 10^{-9} per operating hour	Hp#3
OSF#4	**Provide Safe Passenger Embarkation / Disembarkation**	
OSR4.1	It shall not be possible for passengers to board or leave a moving train	Hp#8, Hp#9
OSR4.2	Except in an emergency, it shall not be possible for passengers to board or leave a stationary train unless the train is in a station and the door through which they embark / disembark is on the side of, and level with a section of, and adjacent to, the edge of the in-use platform	Hp#8, Hp#9
OSR4.3	Measures shall be taken to prevent embarking and disembarking passengers from becoming trapped in closing train doors or platform doors	Hp#9

Requirement ID	Requirement Description	Related EUC Hazard
OSR4.4	Measures shall be taken to prevent embarking and disembarking passengers from falling or becoming trapped between the platform edge and the body of the train	Hp#9
OSR4.5	Minimum dwell times should be maintained so as to allow less mobile or encumbered passengers to leave the train before the doors close	Hp#9
OSR4.6	Measures shall be taken to prevent passengers waiting on a platform being too close to, or falling/jumping from, the platform edge	Hp#12
OSR4.7	Except in an emergency, it shall not be possible for passengers to exit through the ends of a platform	Hp#6
OSF#5	**Provide Safe Maintenance Access to Track**	
OSR5.1	Trains shall be prevented from accessing areas of the railway that must be reserved for maintenance access (i.e. Work Zones)	Hp#9
OSR5.2	It shall be possible to move engineering vehicles into, and out of, Work Zones, *only* with the coordination of those at the worksite	Hp#9
OSR5.3	Maintenance access shall be prevented if trains are running in the proposed Work Zone	Hp#9
OSF#6	**Ensure Guideway is Safe for Train Passage**	
OSR6.1	Obstacles or unauthorised personnel within the swept envelope of the train's route shall be prevented, or shall be detected in time for emergency braking to be applied in order to avoid a collision	Hp#6, Hp#10
OSR6.2	In the event of hazardous damage to track elements or other infrastructure, appropriate action, e.g. temporary speed restrictions in, or closure of, the affected area, shall be taken in order to protect train movements	Hp#10
OSF#7	**Ensure Safe Acceleration & Braking**	
OSR7.1	Except when necessary to respond to a higher-risk situation, sudden / sharp increases or decreases in train acceleration / deceleration (jerking), sufficient to cause injury to passengers, shall be avoided	Hp#4
OSF#8	**Ensure Safety of Traction Power Supply**	
OSR8.1	Trains shall be prevented from feeding a traction power supply section that had been isolated (regenerative train braking)	Hp#11
OSR8.2	Where a traction power supply section had been cut off for on-site maintenance purposes, explicit agreement of those at the worksite shall be required prior to restoration of the supply	Hp#11

It should be noted that these requirements are objective-based (or rule-based) in that they express what the OSFs have to *achieve* rather than what they have to do; this means they form a vital link in the *rich traceability*[14] between the lower-level Safety Functions and the EUC Hazards that the functions are required to mitigate.

The need to specify interim, worst-case success criteria[15] for OSFs 1# to #3, in particular, is based on two related factors:

- the fact that they make the greatest, and most direct, contribution to what the UGTMS standard (IEC 2014b) describes as the "*safe movement of trains*" overall; and
- the reasonable assumption that the processes described in Sub-section 4.4.4 below would lead each of them being assessed as a *SIL 4* function, as defined in IEC 61508.

The key assurance question at this stage is, therefore, whether the above requirements for each OSF would be sufficient to mitigate the corresponding EUC hazard(s) — in other words, are there any conditions (*except* for failures within the OSF or *abnormal* operating conditions) that could lead to the EUC hazard occurring, at an intolerable rate.

Furthermore, the rigour of the assurance required here would depend on the Safety Integrity Level (SIL) for the OSF concerned — see the next Sub-section.

4.4.4 Determine the Safety Integrity Requirements for each Overall Safety Function

According to IEC 61508-1, this step involves the determination of the SIRs required of each of the above OSFs, in order to achieve a tolerable level of risk overall.

IEC 61508-1 states that the SIRs, at this level, must be specified in terms of either:

- the amount of EUC-risk reduction required in order to achieve the tolerable level of risk; or
- the tolerable rate of occurrence of the [EUC] hazardous events, in order to achieve the tolerable level of risk.

There are number of key points to note, as follows.

Firstly, the SIRs at this "overall" level are not, despite their name, properties of the OSF to which they relate[16] — they actually specify a *target* amount of EUC risk reduction that each OSF has to meet[17].

Secondly, it is reasonable to assume that in giving the choice of how to specify the SIRs, IEC 61508-1 intends that the two methods are equivalent, albeit the latter does not require knowledge of what the EUC Risk would have been *before* it was reduced. Therefore, as it is clear that the risk-reduction method depends on the functionality and performance, as well as on the failure rate, of the safety functions, so must the latter method; in other words, though it might be tempting to believe that EUC hazard-occurrence rates could be interpreted directly as OSF failure rates, that would be an entirely false deduction — for further explanation on this point see Fowler (2022).

[14] Traceability that embodies evidence of requirements satisfaction — in this case, evidence that the functional safety properties of Safety Functions are necessary and sufficient to reduce EUC risks to a tolerable level.

[15] For convenience, these are actually expressed a maximum probability of each function being *unsuccessful* in meeting its functional requirements.

[16] Of course, as already seen in Sub-section 4.4.3, this is true also of the *functional* requirements of the OSFs.

[17] Although we will persevere with the IEC 61508 terminology of "safety integrity requirements", such properties might be better thought of as being *safety criteria*, as used in some areas of the transport sector.

Thirdly, notwithstanding the previous point, there is a reasonably straightforward path from knowledge of the most demanding tolerable rate of occurrence of the [EUC] hazardous event, associated with each OSF, to the derivation of a SIL[18] for the OSF (Fowler 2022).

Fourthly, although not a simple mechanical process, methods of deriving tolerable rate of occurrence of the [EUC] hazardous events, from pre-defined *target levels of safety* for the associated accidents, are well documented in rail safety standards such as CENELEC (2017); therefore, since it is not important to the main message of this article to provide a worked example of failure analysis here, we will leave the discussion at this point but pick it up again in the context of lower-level SIRs derivation, in Sub-section 4.6.5 below.

4.5 Overall Safety Requirements Allocation (IEC 61508-1 Phase 5)

4.5.1 Aim

The aim of Phase 5 is to allocate, to Safety Related Systems (SRSs) and/or Other Risk-reduction Methods (ORRMs), the safety requirements, which were derived for the corresponding Overall Safety Functions in Phase 4.

4.5.2 Discussion

IEC 61508 gives prominence to the distinction between SRSs and ORRMs — partly, it would seem because, once identified, the latter measures fall outside the scope of the Standard. Whereas for, say, process industries, the identification of, and distinction between, the two categories of risk-reduction means might be quite straightforward, for the more complex transport applications it is less so.

Table 5 shows a suggested summary allocation of the OSFs from Table 4 on to what might be interpreted generically as SRSs and ORRMs, within the scope of ATC operations. For the purposes of this exercise, the SRSs have been adapted from the top-level functional elements described in the UGTMS standard[19] (IEC 2014b).

Table 5 ~ Allocation of Overall Safety Functions for ATC Operations

OSF ID	OSF Title	SRS(s)	ORRM(s)
OSF#1	Establish & Protect a Safe Route for each Train Movement	SRS#1 - Set & Protect Route Elements SRS#4 - Authorise Train Movement SRS#5 - Supervise Train Movement	
OSF#2	Apply & Maintain Safe Separation between Trains	SRS#2 - Locate Trains SRS#4 - Authorise Train Movement SRS#5 - Supervise Train Movement	

[18] SILs, as defined in IEC 61508-4, are also not properties of an OSF (or of a system, subsystem, element, or component thereof) – see Fowler (2022).

[19] In constructing this table, we noted that the titles of the three most critical top-level safety functions in the UGTMS standard are quite misleading — e.g. "Ensure Safe Separation of Trains" (5.1.2) actually covers only the location of trains — and so we avoided using them, preferring instead to reference the Standard's lower-level functions that addressed the full scope of the OSFs concerned.

OSF ID	OSF Title	SRS(s)	ORRM(s)
OSF#3	Enforce Safe Speed Limits for Trains	SRS#3 - Determine Permitted Speed SRS#4 - Authorise Train Movement SRS#5 - Supervise Train Movement	
OSF#4	Provide Safe Passenger Embarkation / Disembarkation	SRS#6 - Supervise Passenger Transfer	Platform Screen Doors / End Doors, Gap Fillers
OSF#5	Provide Safe Maintenance Access to Track	SRS#7 - Protect Staff on Track	Maintenance Safety Procedures
OSF#6	Ensure Guideway is Safe for Train Passage	SRS#8 - Supervise Guideway	Segregated guideway, fences, walls, bridges / subways, etc.
OSF#7	Ensure Safe Acceleration & Braking	SRS#9 - Drive Train	Train's power and braking systems
OSF#8	Ensure Safety of Traction Power Supply	-	Maintenance & Power Supply Safety Procedures Train's power and braking systems

The ORRMs include mainly non-functional, safety-related items for which separate design and development standards would normally exist but which may be related to the corresponding SRSs.

Further details of the SRSs and ORRMs will emerge during the processes described in Sub-sections 4.6 and 4.7 below, respectively.

4.6 Specification of Safety Requirements for SRSs (IEC 61508-1 Phase 9)

4.6.1 Aim

In IEC 61508, the aim of Phase 9 is to develop safety requirements for the SRSs identified in Phase 5, in terms of their FSRs and SIRs, in order to achieve the required functional safety under all *normal*, *abnormal* and *failure* conditions.

4.6.2 Overview

It is important to note here that IEC 61508-1 places great emphasis on the need for a rigorous description of the workings of SRSs at this level, including:

- a description of all the Safety Functions, how they work together to achieve the required functional safety and whether they operate in low-demand, high-demand or continuous modes of operation;

- the required performance attributes of each Safety Function — e.g. timing properties and, for more data-intensive applications than possibly envisaged by IEC 61508, data accuracy, latency, refresh rate, and overload tolerance;
- all interfaces that are necessary to achieve the required functional safety;
- all relevant modes of *normal* operation of the EUC;
- all other required modes of behaviour of the SRSs — in particular:
    - their required response in the event of defined *abnormal* operating conditions of the EUC or its environment
    - their *failure* behaviour and their required response in the event of such failure (Fowler 2022).

To that end, this Sub-section comprises four stages, as follows:

Firstly, the development of FSRs for scenarios covering the entirety of *normal* operations. This will be done (initially at least) at a relatively abstract level, without any reference to physical elements within the end-to-end ATC system[20] (see Sub-section 4.6.3 below).

Secondly, to show that the FSRs specified for the SRSs would be adequate to meet the risk-reduction required of the SRSs, in the absence of failure (see Sub-section 4.6.4 below).

Thirdly, to analyse, in a similar manner, scenarios covering *abnormal* events in order to identify any additional FSRs necessary to maintain a tolerable level of safety during such events (see Sub-section 4.6.5 below).

Fourthly, to analyse scenarios relating to potential *failures* of the ATC system in order to identify SIRs, and any additional FSRs, necessary to maintain a tolerable level of safety during such failure events (see Sub-section 4.6.6 below).

Because the first three stages are directly relevant to the "IEC 61508 viewpoint", outlined in Sub-section 1.2, and the fourth is addressed in detail in existing railway standards, most of the focus below is on the former stages.

4.6.3 FSRs for Normal ATC Operations

This first stage involves the identification of a set of Safety Functions for each of the SRSs in Sub-section 4.5, and the derivation of detailed functional safety requirements (FSRs) for each Safety Function that, in conjunction with the properties of the associated ORRMs, would ultimately satisfy the OSF requirements of Table 4.

It is evident, especially in the case of a fully automated railway control system, that the high number of Safety Functions (and an even-higher number of associated detailed FSRs) would be very large. Fortunately, that task is made very much less daunting by the publication, in IEC (2014b), of a comprehensive, generic functional requirements specification for UGTMS, which we can use as a starting point for our urban railway example, as set out initially in Table 6.

Table 6 shows, for each SRS derived in Table 5, a description of the Safety Functions that make up that SRS. It should be noted that these Safety Functions are limited to those that are necessary to address *normal* ATC operations and might not be sufficient for the system to specific how safely ATC must react to *abnormal* operating conditions (Sub-section

[20] As noted in Sub-section 3.7.2 of Fowler (2022), the IEC 61508 objective here is to "*describe, in terms not specific to the equipment, the required safety properties of the SRS(s)*". This level of requirements expression respects that objective since it makes no assumptions about the technology involved in the realisation of the requirements.

4.6.5 below) or to provide mitigation of ATC system internal *failures* (Sub-section 4.6.6 below).

Table 6 ~ ATC Safety Functions per Safety Related System

SRS	SF ID	Safety Functions	Description
Set & Protect Route Elements (SRS#1)	SF#1.1	Reserve, Set & Lock a Route	Establishes (i.e. reserves, sets & locks) a standard route in response to a route call.
	SF#1.2	Supervise Route	Supervises that all conditions for the route are still in place.
	SF#1.3	Maintain Route Locking	Keeps the route locked against route release by manual or system input: • for an approaching train for which the movement authority allows entry into the route, or • for a train that is already within the route.
	SF#1.4	Release Route	Releases the route when all of the conditions for maintaining it locked no longer apply
Locate Trains (SRS#2)	SF#2.1	Initialise Reporting Trains Location	Initialises location of reporting trains which are: • stationary in stabling locations, • entering ATC territory, • recovering from localisation failures.
	SF#2.2	Determine Train Orientation	Determines physical orientation of train relative to defined orientation of the track.
	SF#2.3	Determine Train-travel Direction	Determines the actual travel direction of reporting trains, relative to the track.
	SF#2.4	Determine Reporting Train Location	Determines the location of all reporting trains according to the train orientation and train length.
	SF#2.5	Determine Non-reporting Train Location	Determines if a section of track is occupied by non-reporting trains based on inputs received from devices external to the ATC system.
Determine Permitted Speed (SRS#3)	SF#3.1	Determine Permanent Infrastructure Speed Profile	Determines the permanent speed profiles, based on infrastructure data, e.g. track geometry & quality, and infrastructure constraints (tunnels, bridges, platforms, etc.).
	SF#3.2	Determine Temporary Infrastructure Speed Restrictions	Sets and removes temporary speed restrictions for selected areas by operational commands or as result of system reactions.

SRS	SF ID	Safety Functions	Description
	SF#3.3	Determine Permanent Rolling Stock Speed Restrictions	Determines the maximum permitted speed for each type of rolling stock.
	SF#3.4	Determine Temporary Rolling Stock Speed Restrictions	Determines temporary rolling stock speed restrictions due to train failures and to driving modes.
Authorise Train Movement (SRS#4)	SF#4.1	Determine Limit of Movement Authority	Determines for each train its limit of the movement authority (LMA), corresponding to the first conflict point ahead of the train.
	SF#4.2	Determine Train Protection Profile	Determines the train protection profile for all trains to ensure their LMAs and authorised speeds are never exceeded.
	SF#4.3	Authorise Reporting Train Movement	Authorises train movement for reporting trains in accordance with its Train Protection Profile.
	SF#4.4	Authorise Non-reporting Train Movement	Authorises train movement by wayside signals if conditions of safe route and safe separation are fulfilled.
Supervise Train Movement (SRS#5)	SF#5.1	Determine actual train speed	Determines the actual train speed.
	SF#5.2	Supervise Safe Train Speed	Supervises actual train speed against the permitted speed with respect to the Train Protection Profile.
	SF#5.3	Supervise Safe Train Direction	Supervises movement of a train against the authorised direction of travel.
	SF#5.4	Supervise Movement-Authority Validity	Monitors validity of a train's movement authority and determines action to be taken if validity period is exceeded.
	SF#5.5	Overrun Protection	Supervises the actual position of a train against its LMA.
Supervise Passenger Transfer (SRS#6)	SF#6.1	Control Train & Platform Doors	Contains functions and requirements that are able to authorise and command the opening and closing of train doors, and platform doors, once all conditions which are required to ensure a safe passenger transfer have been met.

SRS	SF ID	Safety Functions	Description
	SF#6.2	Prevent Injury to Person between Train and Platform	Controls external devices and supervises detectors that prevent injuries to persons from falling (or detect persons falling) and becoming trapped between the platform edge and the train body.
	SF#6.3	Authorise Safe Station Departure	Authorises the train to leave the station only when all train doors and all platform doors) are closed and locked.
Protect Staff on Track (SRS#7)	SF#7.1	Protect staff on track by Work Zone	Establishes, and subsequently removes, Work Zones in order to protect staff on the track.
Supervise Guideway (SRS#8)	SF#8.1	Prevent collision with obstacles	Contains functions and requirements that are able to prevent, or detect, collisions with obstacles present in the guideway.
	SF#8.2	Prevent collisions with persons on tracks	Contains functions and requirements that are able to prevent collisions with persons who mainly could enter from platforms to track areas.
Drive Train (SRS#9)	SF#9.1	Determine Operating-speed Profile	Determines the Operating Speed Profile, taking into account ride quality, passenger comfort and the driving mode, (including service acceleration/deceleration rate), within the constraints of the Train Protection Profile.
	SF#9.2	Control Train Movement	Determines, and sends to the rolling stock, traction and braking commands to ensure that the train speed follows the train operating profile and to achieve accurate stopping.

From this point on, we run into a potential problem of developing far too much detail for this article to handle; e.g. for the 30 Safety Functions shown in Table 6, there is a total of around 350 associated Functional Safety Requirements (FSRs)! Therefore, in the illustration at Appendix B, we have shown only the Safety Functions / FSRs that apply to the three SRSs that are needed to support the overall safety function OSF#1, "*Establish & Protect a Safe Route for each Train Movement*".

4.6.4 Adequacy of the Functional Safety Requirements

Thus far, the SRSs, Safety Functions, and their safety requirements, have been derived purely hierarchically, and what we have yet to show explicitly is, *inter alia*:

- how the functions interact with each other, and with the elements of the wider ATC system and (what we described, in Sub-section 4.1.4 as) the EUC Control System;
- the information needed by, and produced by the SRSs and Safety Functions;
- the system states and sequence of events, during a typical "day-in-the-life" of a train;
- any additional functionality to cope with abnormal and failure events; and
- whether the requirements constitute a complete, correct and coherent set.

There is a wide range of techniques for addressing these issues, and the following are examples of a few of them.

State-transition Models (STMs): a simplified example of which is shown in Figure 4, are used to capture knowledge about system behaviour. They can be translated into other models to support qualitative analysis (e.g. Sequence Diagrams and Activity Diagrams) or quantitative analysis (e.g. Markov models). They represent system behaviour in the form of: the state space of the system in a given context; the events that cause a change of state; the transitions between states; and resulting actions (Lucic 2015).

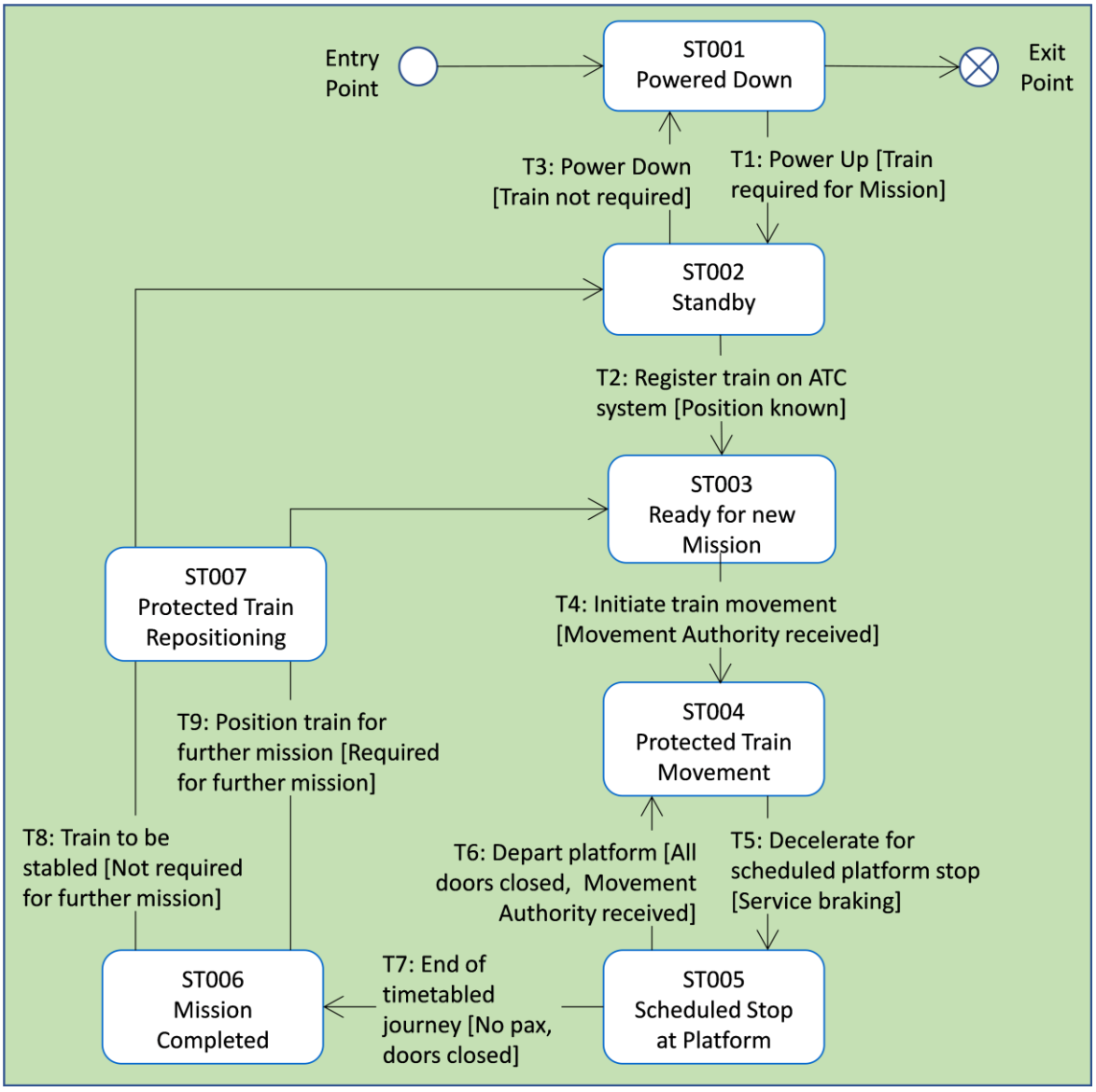

Figure 4 ~ State-transition Diagram for *Normal* Operations

The specific context for Figure 4 is our ATC-enabled railway, seen from the perspective of an ATC-capable train, for the whole of a typical day, i.e. under *normal* operating conditions[21]. The seven states are shown as rounded rectangles, and the nine permitted transitions between the states are represented by the arrows, which are accompanied by

[21] A more complete model would need to include, for example, states applicable to non-ATC trains, Engineering Hours & Possessions, and *abnormal* & *failure* conditions.

text signifying the trigger[22] for, and (in square brackets) any conditions or constraints[23] applicable to, the transition.

The initial state is the train powered down in a depot (ST001). When a requirement for the train to undertake a mission[24] is imminent, the train is brought to a *standby* state (ST002) in which all its systems are running as required but the train is not yet *registered* on the wider ATC system.

When the train's position is known, the train will be registered automatically leading to a state of readiness for undertaking a new mission (ST003).

Once the train has received a Movement Authority, it will move as required by its timetable, under the protection of the ATC system (ST004), until it stops at the first scheduled platform and the appropriate doors are opened (ST005). The train will then repeat states ST004 and ST005 until it reaches its final scheduled stop, all passengers having disembarked and the doors closed (ST006).

Transition T9 then provides for the train to be repositioned (ST007) to undertake further missions until ST006 applies to the final mission of the day. Finally, transition T8 provides for the train to return safely to depot at the end of the day (ST007).

We have chosen to use STMs at this high level in the system hierarchy in order to provide an overarching framework for the next level of analysis, which uses Sequence Diagrams.

Sequence Diagrams: an example of which is shown in Figure 5, is a dynamic form of interaction diagram that shows *objects* (and / or actors) whose *lifelines* run down the page, and with the interactions between them represented as a sequence of messages that are drawn as arrows from the source lifeline to the target lifeline (Sparx Systems 2022).

In this context, we use such diagrams as a method for describing *operational scenarios*, which can be thought of as:

> *"A set of actions or functions representing the dynamic of exchanges between the functions allowing the system to achieve a mission or a service". (SEBoK 2022).*

The Sequence Diagram shown in Figure 5 is for a specific operational scenario in which an ATC-capable train makes a protected journey between two stations, behind a non-ATC train.

At this level of analysis, we have chosen the functional objects to be SRSs (in blue)[25] implying that blue lifelines represent Safety Functions. There are also three actors: which represent the two basic trains[26] and various wayside devices (including demandable elements and occupancy-detection components). One of the great strengths of this technique is that, in later phases of the lifecycle, the actors and objects of the same scenarios can be redefined, at other levels, e.g. logical design, physical design and software module levels; thus, as SEBoK (2022) also notes:

> *"Operational scenarios are used to evaluate the requirements and design of the system and to verify and validate the system".*

[22] An event or action.

[23] Also known as "guards".

[24] A mission is *planned* journey of the train between two fixed (start and destination) points including any scheduled stops.

[25] For completeness, the SRS "Supervise Train Movement" should also have been included but has been omitted for the sake of simplicity of the diagram

[26] "Basic" means that train-borne ATC elements are not included, and is consistent with the four ATC *objects* being purely functional

In this scenario, there are:

- two trains: Train 1 is an ATC-capable reporting train and is following Train 2, which is a non-ATC / non-reporting train;
- three contiguous routes, A, B and C, in an interoperability area in which reporting trains and non-reporting trains move, under ATC control (the latter trains being controlled via wayside signals);
- two stations: the first in Route A, and the second in Route B.

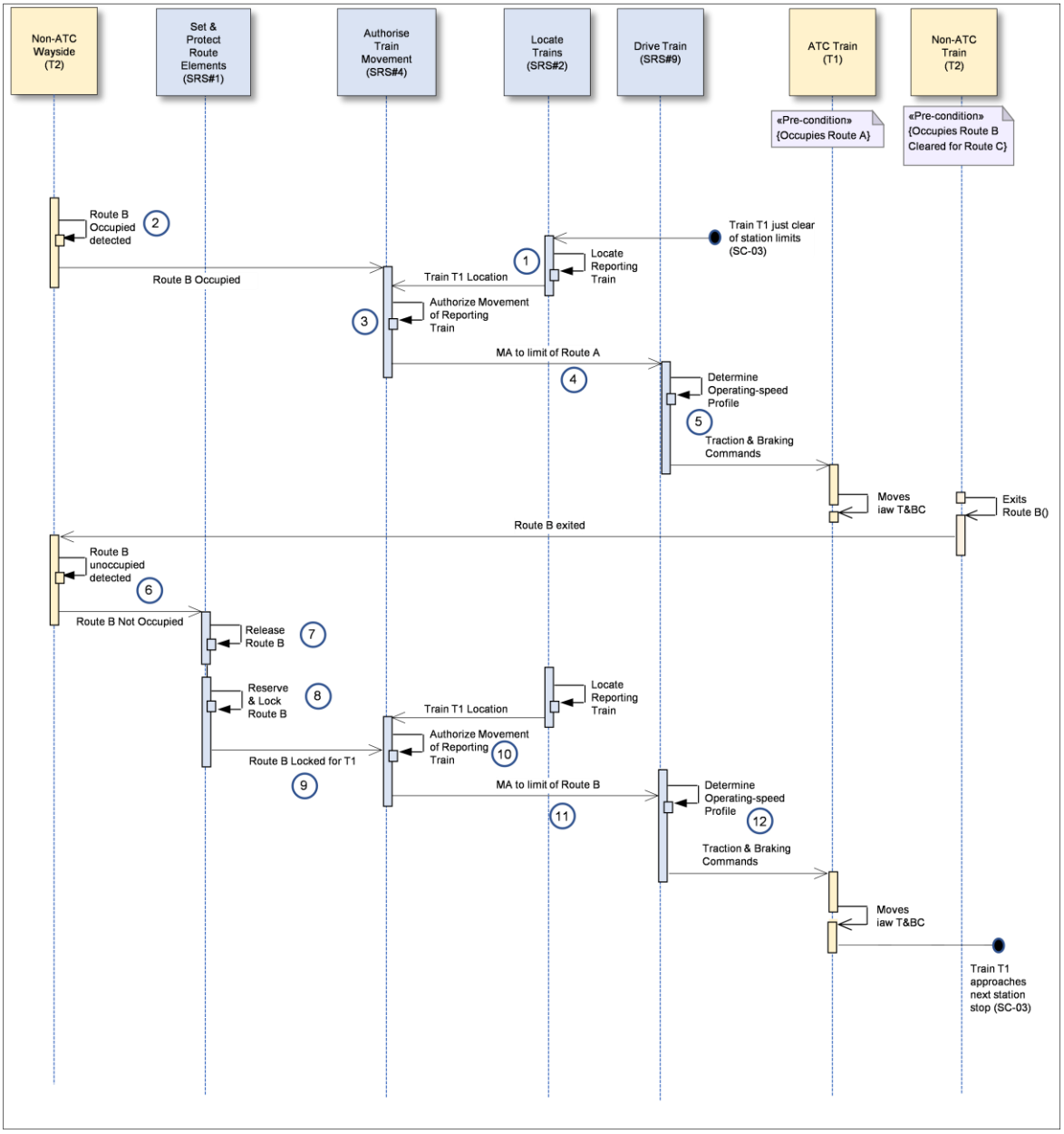

Figure 5 ~ Sequence Diagram for a Protected Movement Between 2 Stations (SC-02)

At the start of the scenario:

- Train 1 is in Route A and leaving the first station (the end of scenario SC-03[27], "*ATC Train Makes Scheduled Station Stop*"), and heading for the second station;

[27] See Appendix C

- Train 2 is in Route B, having departed the second station and has already been cleared for Route C.

Table 7 provides an outline narrative of the subsequent events that are numbered #1 to #12 on the diagram, together with a reference to the related Safety Functions.

Table 7 ~ Scenario SC-02 Narrative of Events

#	**Description**
1	Location of Train 1 in Route A is determined and reported (SF# 2.2. to 2.4) to *Authorise Train Movement*
2	Non-ATC Wayside detects , and reports, Route B as being occupied (SF# 2.5) by Train 2
3	*Authorise Train Movement* maintains the LMA for Train 1 at the safe limit of Route A (SF# 4.1. to 4.2)
4	*Authorise Train Movement* sends Movement Authority (SF# 4.3) to *Drive Train* accordingly
5	*Drive Train* determines the *Operating- speed Profile* (SF# 9.1) for Train 1 and issues traction / braking commands (SF# 9.2), to *Train T1*, for the train to continue to move in accordance with the *Operating- speed Profile*
6	Meanwhile, Train 2 exits Route B; non-ATC Wayside detects this (SF# 2.5) and reports, to *Set& Protect Route Elements*, that Route B is "not occupied"
7	*Set & Protect Route Elements* releases the Route B accordingly (SF# 1.4)
8	When Route B becomes available, *Set& Protect Route Elements* sets and locks that route (SF# 1.1 to 1.3) for Train 1.
9	*Set & Protect Route Elements* reports, to *Authorise Train Movement*, that Route B has been set and locked for Train 1
10	*Authorise Train Movement* updates the LMA for Train 1 to be at the safe limit of Route B (SF# 4.1. to 4.2)
11	*Authorise Train Movement* sends the new MA (SF# 4.3) to *Drive Train* accordingly
12	*Drive Train* determines the *Operating-speed Profile* (SF# 9.1) for Train 1 and issues traction / braking commands (SF# 9.2), to *Train T1*, for the train to continue to move in accordance with its *Operating-speed Profile* until the train approaches the next station stop (beginning of SC-03).

In effect, the Sequence Diagram for scenario SC-02 details, at a safety-function level, the interactions between the ATC SRSs in respect of state ST005 in Figure 4, as entered via trigger T9 and exited via T8. This demonstrates the role that an STM can play in deriving a complete set of Operational Scenarios for a given context; Table 10, at Appendix C hereto, gives examples of some of the Operational Scenarios that would be needed to underpin the full range of states and transitions shown on Figure 4. That said, a considerable amount of operational and technical expertise and effort would be needed to ensure a complete, correct and coherent set of Operational Scenarios is derived and analysed in practice.

It would be very important at this stage to crosscheck also the diagram and narrative against the detailed FSRs for each Safety Function, to ensure that the completeness and correctness of the relevant set of FSRs in each case.

Since the scenario analysis tells us much more about the required system dynamic behaviour than the purely textural FSRs, each scenario description should also be considered to be an FSR in its own right.

What the Sequence Diagram does not capture is the full complexity of, interactions between, and data used /produced by, each of the Safety Functions that constitute each SRS; for this, we can use Activity Diagrams.

Activity Diagrams: an example of which is shown in Figure 6, are essentially an advanced version of flow charts that model the flow from one activity to another activity.

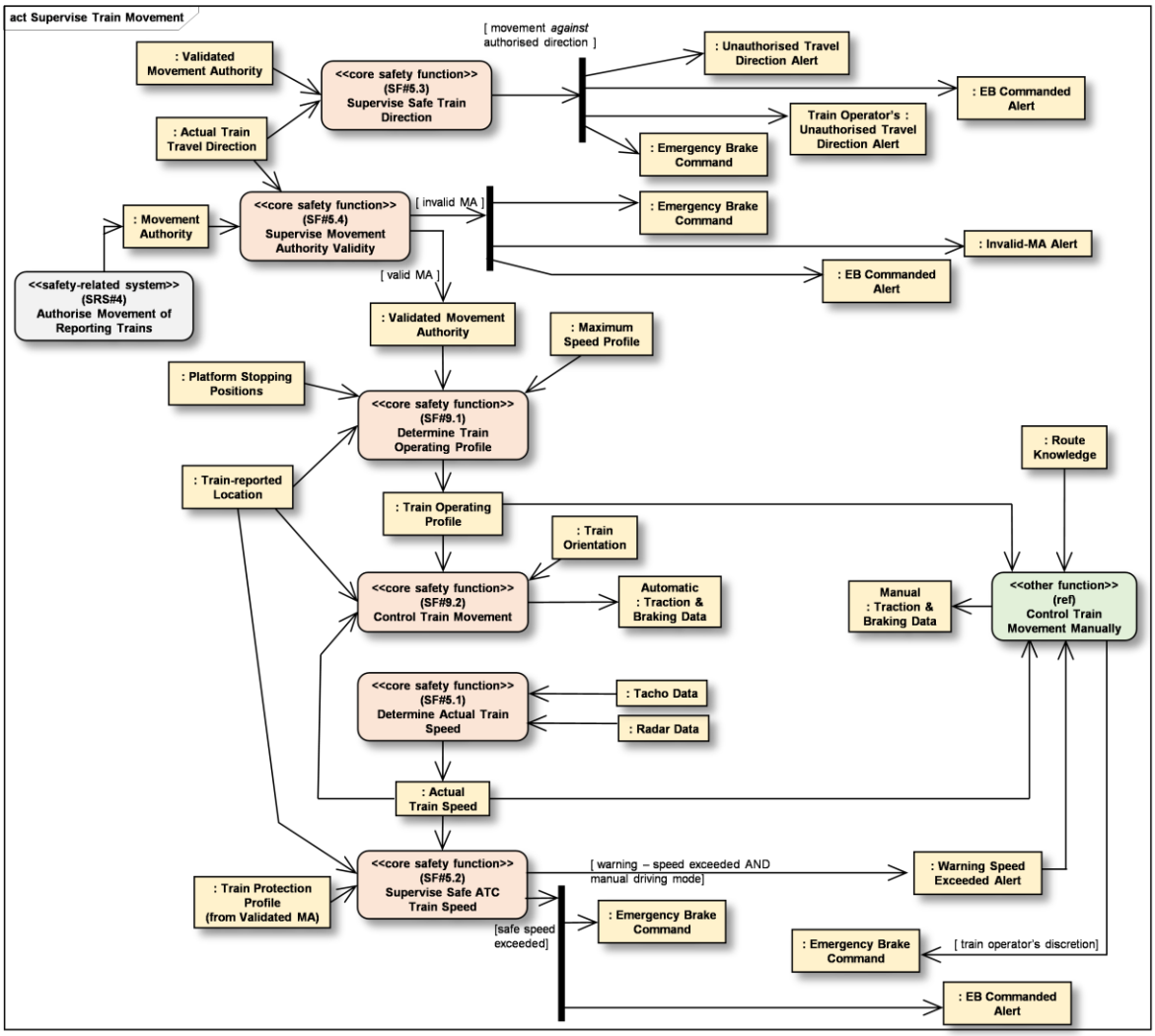

Figure 6 ~ Activity Diagram for Supervise Train Movement

This activity diagram is for the SRS "*Supervise Train Movement*", which supports each of the first three (and most safety-critical) of the Overall Safety Functions shown in Table 5. The diagram in this case is more structural than sequential, since the overall process is iterative and some of the functions might be running concurrently or asynchronously. The closely related SRS "*Authorise Train Movement*" is also shown in outline, for reference.

The activities (rounded rectangles) take the form of the Safety Functions involved, with rectangles representing the associated data, i.e. the information produced or used by the Safety Functions, and the instructions that they are required to issue or react to. The arrows indicate the required direction of flow of that data. Two possible modes of operation are covered: i.e. the train being driven automatically, or manually.

Since it would normally be impractical for purely textural FSRs to capture all the information presented in an Activity Diagram, the diagram itself should be identified as a safety requirement in its own right.

Given that we have already identified more than 30 Safety Functions, across nine SRSs, the use of a software design tool would not just help in the diagram's construction, it would also play a crucial part in in preserving the uniqueness of the functions and data, both within the full set of Activity Diagrams (covering, for example, all SRSs) and between those diagrams and the Sequence Diagrams discussed above.

Accepting that the need for some additional FSRs (or even Safety Functions) would probably be identified subsequently, in order to mitigate the effects of specific *normal*, *abnormal* and *internal-failure* events, the functional model of the ATC system, i.e. the aggregate of the Activity Diagrams, once determined, remains sensibly constant; unlike Sequence diagrams, which are very much context dependent.

Other techniques: there are many other techniques that can be used to model the system at this, and lower, levels of representation; a review of some such techniques is presented by Lucic (2015).

4.6.5 ATC Operations under Abnormal Operating Conditions

In general, *abnormal* conditions stem from two main sources:

- hazardous events in the operational environment that are not encountered on a day-to-day basis — hazards Hz#13 to 17, in Table 2, are good examples of such events; and
- failure events within the EUC Control System but *outside* the scope of the ATC system itself, e.g. a failure of a train's traction-control system.

In either case, what we are interested in, first of all, are the following:

- what effect the event would have on the continuing functioning of the ATC system and the consequences for the safe operation of the railway; and
- what actions would need to be taken to mitigate the consequences of the event, and how the functionality of the ATC system (existing or additional) could be used to support such actions.

One very useful way of modelling such events is through Operational Scenarios, based on Sequence Diagrams, as described (for *normal* operations) in Sub-section 4.6.4. First of all though, we need the equivalent of the State Transition Diagram of Figure 4, but covering *abnormal* states; this is shown in Figure 7 overleaf.

State ST008 simply provides a link to / from normal operations; the four main abnormal states are then as follows:

- ST009, Degraded Operations: a sub-optimal operational state of the railway where the train/service is able to continue with its mission, or a state where a fault or a combination of faults and external circumstances results in inability to continue under planned operation;
- ST010, Emergency Operations: response of the service/system to a hazardous event, usually external to the EUC Control System and ATC System, which requires immediate action;
- ST011, Recovery: process of returning the system to an operational state, and recovering the service to its planned operational state, following an emergency, or

degraded operation — may include rescue, involving either rescuing a train, rescuing passengers from the train, or both; and

- ST012, Mission Aborted: a mission for a train(s) has been terminated, following unacceptable circumstances.

From this analysis, Table 11, at Appendix C hereto, gives examples of some of the Operational Scenarios that would be needed to cover the full range of *abnormal* conditions.

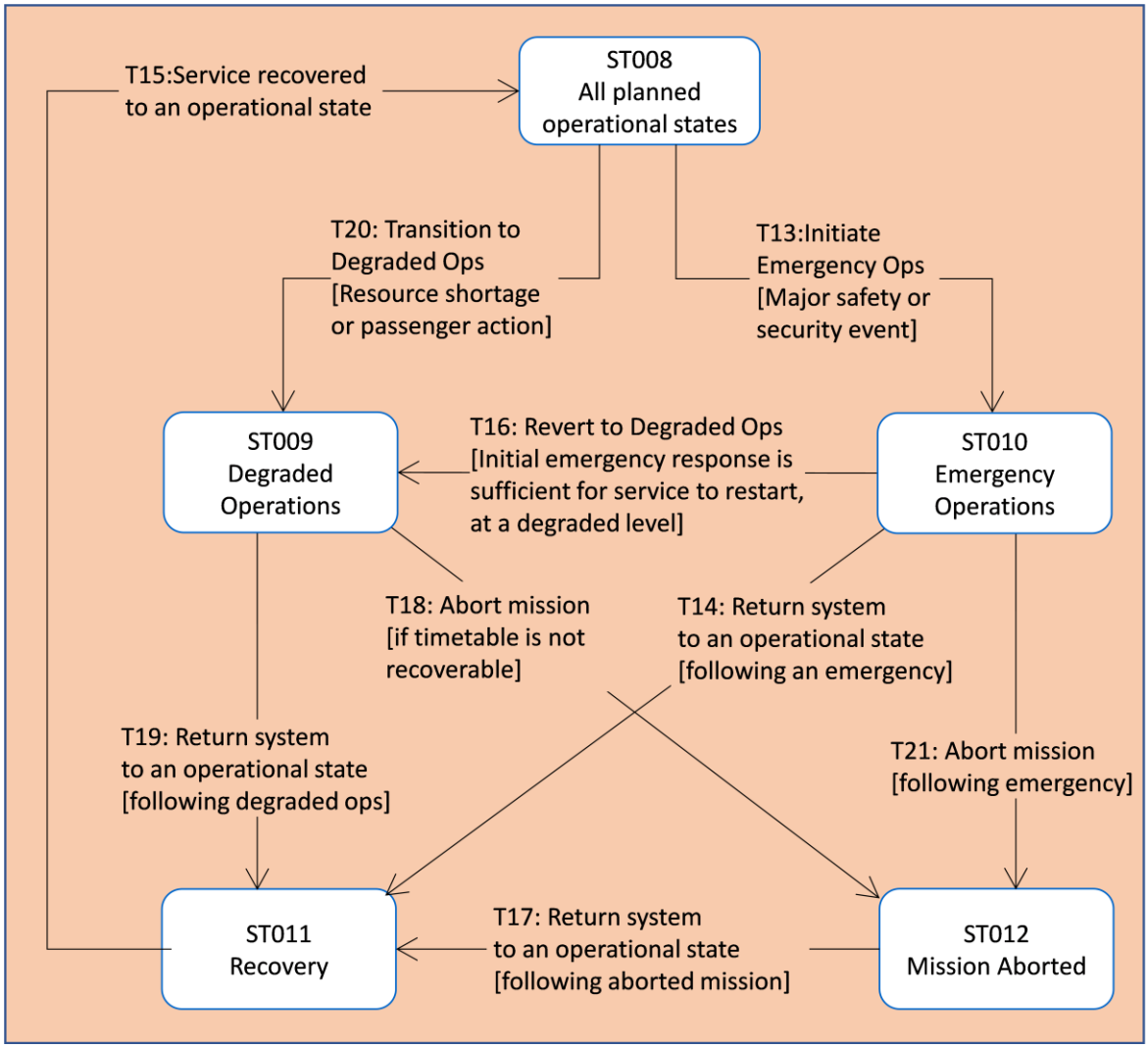

Figure 7 ~ State Transition Diagram for *Abnormal* Conditions

The final step would be to assess any additional risk that would be presented by the event, based on how effective the mitigating actions would be and how often the event is likely to occur. Therefore, in terms of the main safety requirements that might result from the analysis, the most likely would be for operational procedures and/or new *functional* safety requirements for the ATC system. This is unlike the case of failures *internal* to the ATC system, where a major, *additional* output would be safety *integrity* requirements for the ATC system, as we will see in the next Sub-section.

4.6.6 ATC Operations under Internal-failure Conditions

Finally, for Phase 9, is the specification of Safety Integrity Requirements (SIRs) for the SRSs and their Safety Functions, through analysis of potential failures internal to the ATC system. However, given that such analysis is covered comprehensively in existing rail safety standards, including EN 50126-1 (CENELEC 2017), this Sub-section is limited to addressing key principles relating specifically to the IEC 61508 viewpoint.

It is acknowledged that deriving a true "risk picture" for particular operational applications is far from easy and, in keeping with Sub-section 1.1, the following method is offered simply as a suggested approach to solving that problem.

Figure 1 in Sub-section 1.2 shows, for a single system-safety element, e.g. an OSF, a graphical representation of the relationship between its safety properties and their effect on achievable risk, according to IEC 61508. This graph is now presented in the form of a Fault Tree in Figure 8 below[28]; in this simplified example, we see how OSF#3, "*Enforce Safe Speed Limits for Trains*", in effect acts as a potential *barrier* to the EUC hazardous event progressing through to an accident — in this case, a derailment[29].

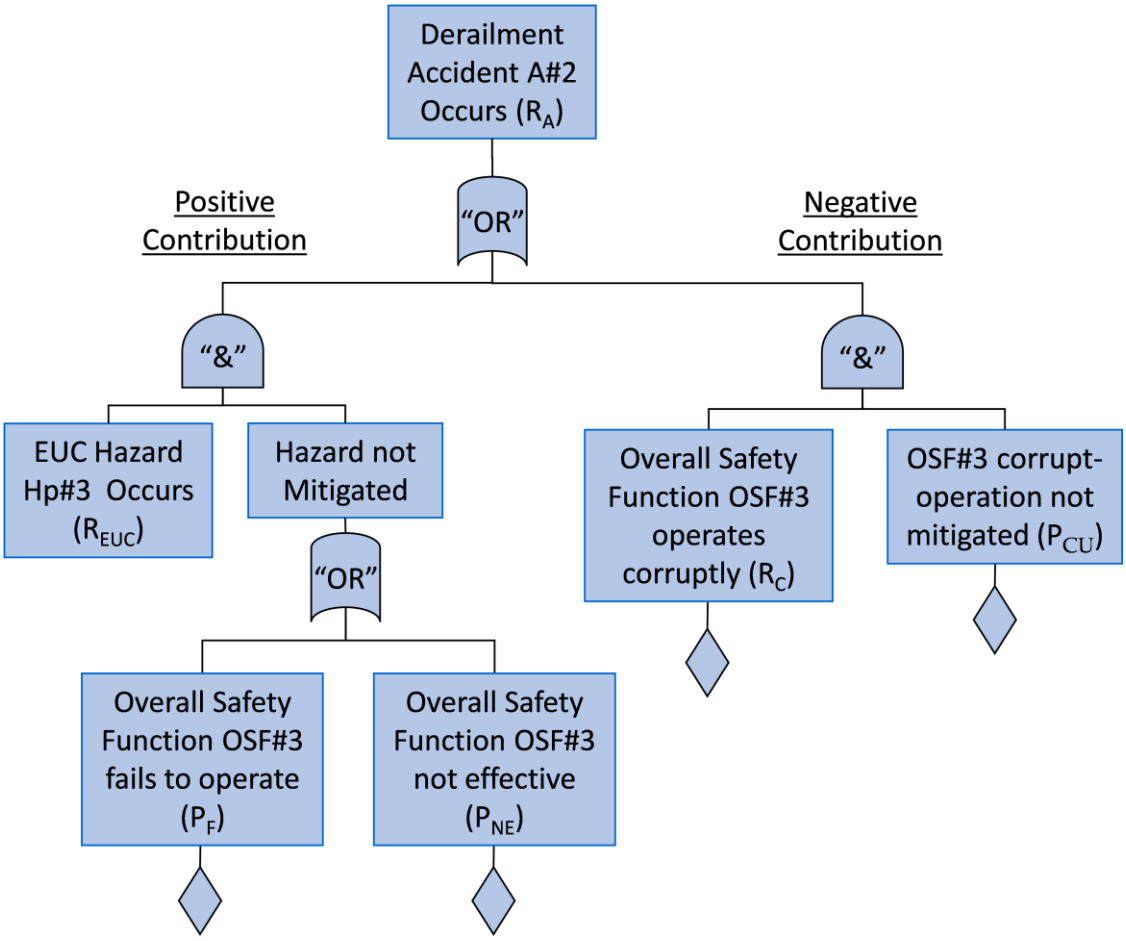

Figure 8 ~ Fault Tree View

[28] In mathematical terms, the fault tree applies to a safety function, in what IEC 61508-4 defines as "*a low demand mode of operation*". However, the authors' intention here is not to detail a quantitative approach — rather, it is to present the general relationships involved.

[29] In this simple example, we have not explicitly captured the possibility of a near miss, i.e. the likelihood that a collision accident would not result even if a hazard was not mitigated, simply because of the 'geometry' of the situation, for example. It is, however, addressed in the subsequent discussion on the Barrier Model.

As we saw in Sub-section 1.2 above, whether or not OSF#3 mitigates the *consequences* of an EUC hazardous event (Hp#3) would depend on its effectiveness when working ($1-P_{NE}$), and the probability that it doesn't fail to operate ($1-P_F$). Although corrupt operation of the OSF could itself lead to a hazardous situation, the rate at which such failures might occur would always be dominated by events outside of the ATC system, i.e. failures in the EUC Control System, or abnormal conditions in the environment.

That said, there might also be means of *reducing* the rate at which an EUC hazard occurs in the first place and to illustrate this we can go to the top-level view of the ATC, described in Section 3, and set them out in the form of a Barrier Model, as in Figure 9.

On the left of the diagram is the input of unmitigated EUC Hazards that are *inherent* in railway operations. Each Barrier, acting in rough sequence from left to right, effectively "filters out" a proportion of the EUC hazards, either by removing them or mitigating their consequences. The safety contribution of ATR is less obvious than that of ATS or ATP, but the argument is that a well-designed and well-run train timetable would (for good business reasons, if nothing else) reduce congestion and, therefore, reduce the number of *opportunities* for a collision accident to occur.

The three main barriers are supported by safety functions or management functions, which are themselves implemented in the physical ATC system, comprising people, equipment and procedures. Of course, these system elements can fail to operate, effectively reducing the probability of success of the barrier, or operate corruptly, giving rise to new, *system-generated* hazards.

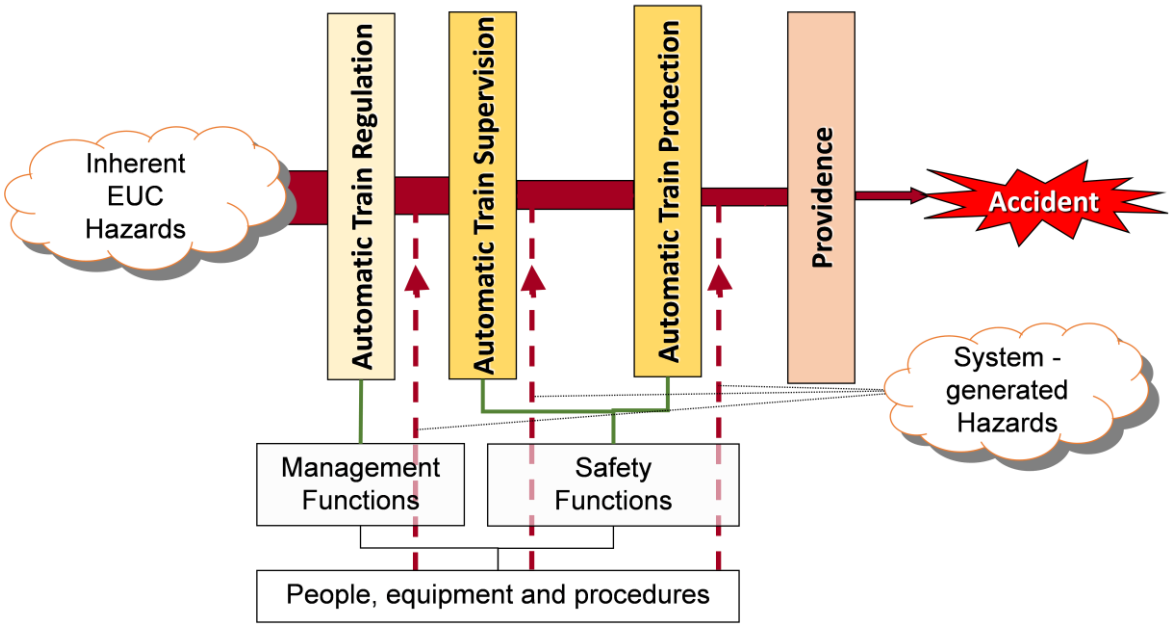

Figure 9 ~ Top-level Rail System Barrier Model

The final, Providence, barrier reflects the point that, even when all three layers of ATC have been unable to remove a hazard, there might still be a significant probability that an actual accident would not result.

In order to quantify the relationships involved, the Barrier Model can be presented in the form of the simplified, top-level Event Tree shown in Figure 10.

At the input to the tree are the unmitigated EUC hazards, which are inherent in railway operations, and which occur at frequency F_U; each barrier then has a probability of

success (P_{Sn}) in mitigating the hazards at its input node (shown thus ⊗), enabling the computation of the risk of an accident (R_A) as:

$$R_A = F_U \cdot (1-P_{S1}) \cdot (1-P_{S2}) \cdot (1-P_{S3}) \cdot (1-P_{S4}) \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

The frequency of other, more-benign outcomes can be similarly computed, with the model capturing the net positive, as well as the negative, contributions of each barrier to the safety of railway operations, in line with the IEC 61508 viewpoint.

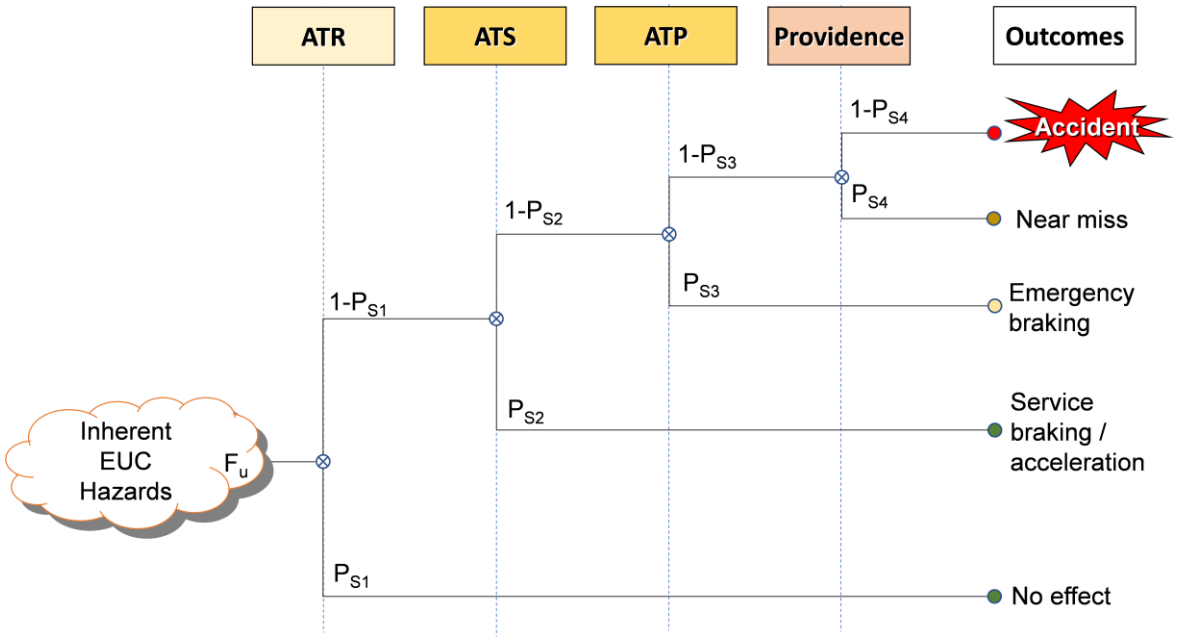

Figure 10 ~ Top-level Rail System Event-tree Model

Of course, the model, as presented here, is purely illustrative and very high-level. Nevertheless it provides a sound framework, for each accident type, and has the advantage of being able to capture multiple *end events* — unlike the Fault Tree, which has only one.

This could be done by developing such an Event Tree for each appropriate accident type of Table 1 and, for each tree:

- decomposing each Barrier into its constituent OSFs, SRSs and Safety Functions;
- constructing Fault Trees (of the form of that shown in Figure 8, but decomposed down to SRS or Safety Function level); and
- linking the Fault Trees to the relevant nodes of the Event Tree such that the top-level event in the Fault Tree represents the probability of a success outcome (P_{Sn}) for the Barrier concerned.

That, of course, raises the more difficult question as to how to get realistic estimates of the probability values for, and a sensible balance between, the Barriers. Fowler and Fota (2023) outlined how this problem had already been addressed and largely resolved, in the Air Traffic Management (ATM) sector, on the European Commission's Single European Sky ATM Research (SESAR) programme, and from which the approach outlined above was derived.

In seeking to overcome many of the shortcomings of more traditional failure-analysis techniques, e.g. hazard-severity / risk-classification schemes, discussed in Fowler (2022), the SESAR approach:

- uses real accident and incident data to populate the model with the historic probability and frequency values;
- more-accurately captures the progression of a hazardous event through to an accident;
- is capable of modelling the interdependencies between barriers, including lower-level common-cause and common-mode failures, which are implied in Figure 9;
- can be adapted so that they properly reflect the operational environment for specific applications;
- is capable of being modified so that the effects, on the historic risk picture, of the introduction of changes at the operational and/or technological level (e.g. the introduction of a new railway control system), could be assessed and, thereby, new risk models produced.

In practice, the SESAR models are used to generate easy-to-use Risk Classification Schemes that more-realistically reflect the overall safety-risk picture of the operational environment concerned.

4.7 Specification of Safety Requirements for ORRMs (IEC 61508-1 Phase 10)

As noted earlier, it has been assumed that all Other Risk Reduction Measures would already exist as part of the legacy railway infrastructure. Therefore, the specification of requirements for these items is not appropriate for such items, in this case.

5 Conclusions

Fundamentally, the IEC 61508-1 lifecycle, as outlined in the first of three articles (Fowler 2022), stems from the simple concept that where there exists an inherently hazardous *Equipment Under Control* (EUC), which presents an intolerable level of risk to its environment, so there is a need to develop and deploy *Safety Related Systems* (SRSs), and/or *Other Risk-reduction Measures* (ORRMs), in order to *reduce* that risk to a tolerable level.

True to its pan-industrial principles, IEC 61508 allows for an EUC to be anything from a nuclear reactor or a chemical process, to road traffic flows, the flow of aircraft through a block of airspace (as in Fowler and Fota (2023)), or the movement of trains around a rail network (as in this article).

Self-evidently, it is the *functional* safety properties of the SRSs / ORRMs that determine their potential to *reduce* the risks, inherent in the EUC, to a tolerable level. Only then does it make sense to consider the safety *integrity* properties of the SRSs / ORRMs, which negatively affect EUC risk in two possible ways:

- loss of function of the SRSs / ORRMs, which would lower the amount by which *inherent* EUC risk could otherwise be reduced;
- corrupt / spurious operation of the SRSs / ORRMs, which would introduce *new* EUC hazards and risks.

This is exactly what the IEC 61508-1 lifecycle does, although the scope of all three articles was limited to the seven IEC 61508 lifecycle phases that relate to the specification of safety requirements, because most of the key principles underpinning IEC 61508 take effect during these earlier phases.

The example application, herein, of the IEC 61508-1 lifecycle to a new, moving-block Automatic Train Control system, for a hypothetical Metro, has:

- provided a comprehensive set of EUC hazards, inherent to rail operations in general;
- presented a systematic way to analyse a system that results in the description of an exemplary set of SRSs[30] and a detailed specification of their constituent Safety Functions, which are required in order to provide the necessary *reduction* in the EUC risk; and
- outlined an effective method of modelling the effects of failure of the Safety Functions such that safety integrity requirements for those Safety Functions could be derived.

That said, the challenge of demonstrating correctness and completeness of the safety analysis processes should not be underestimated since, as past experience suggests, we would probably be dealing with "SIL-4" functions in most cases. A complete response to that challenge is beyond the scope of this article but the following two areas of the process presented herein go some way towards meeting it.

The first is the role of the IEC 61508 concept of "Overall Safety Functions" (OSFs), which at first seemed to be somewhat redundant but soon proved to be a vital link between the EUC hazards and the SRRs / ORRMs that are required to mitigate them. It was already realised (Fowler 2022) that the safety integrity requirements at that level are not actually properties of (but are targets to be met by) the OSFs and, by applying the same logic, we realised the need for a functional equivalent, in the form of *rules-based* requirements.

The second is the use of a hierarchical set of models that capture the required behaviour of, and interactions between, the SRSs and their Safety Functions. Not only do these prove to add an essential dynamic dimension to the rather static individual functional specifications, but they also helped identify missing, incorrect and missing requirements.

Overall, it is concluded that following the principles of the specific phases of IEC 61508 provides a considerable overall benefit of ensuring a top-down, and far more complete, approach to functional-safety assessment than might otherwise be the case. Fowler (2015) observed, *inter alia*, that European rail safety standards at *that* time were based almost entirely on a bottom-up analysis of the risks from failure of safety functions and a tacit (and totally unjustified) assumption that a tolerably safe state of a rail control system would exist provided the system were sufficiently reliable. What we believe has yet to be properly demonstrated (i.e. not merely asserted) is that the *current* set of European rail safety standards do not suffer from the same deficiencies!

Acknowledgments

The authors wish to acknowledge the considerable help, support and understanding of many colleagues (past and present) from Transport for London and beyond, without which this article would not have come to fruition. Our particular thanks go to Dr Ivan Lucic for his guidance and allowing us to make use of material from his book "*Risk and Safety in Engineering Processes*".

The copyright holder of the quotations from published standards used for illustration in the main body of this article is the International Electrotechnical Commission, Geneva; and that of the quotation in Sub-section 1.1 is CENELEC, the European Committee for Electrotechnical Standardization, Brussels.

[30] No new ORRMs were identified in this case.

References

CENELEC. (2017) *Railway applications – the specification and demonstration of reliability, availability, maintainability and safety (RAMS), Part 1: Basic requirements and generic process*, EN 50126-1:2017. European Committee for Electrotechnical Standardization, Brussels.

Fowler D. (2015). *Functional Safety by Design — Magic or Logic?* In Proceedings of the 23rd Safety-Critical Systems Symposium, Bristol, UK. Available at https://scsc.uk/r129/7:1. Accessed 19th June 2022.

Fowler D. (2022). *IEC 61508 Viewpoint on System Safety in the Transport Sector: Part 1 — An Overview of IEC 61508*, in Safety-Critical Systems eJournal, Vol. 1, Iss. 2. Available at https://scsc.uk/r176.3:1, Accessed 29th December 2022.

Fowler D. and Fota O.N. (2023). *Safety Assessment of Point Merge Operations in Terminal Airspace — an IEC 61508 Viewpoint*, in Safety Critical Club eJournal, Vol 1, Iss. 3. Available at https://scsc.uk/r183.3:1#page=25, Accessed 17th July 2023.

IEC. (2010). *Functional Safety of Electrical/electronic/programmable electronic Safety-related Systems*, IEC 61508, V 2.0. International Electrotechnical Commission. Geneva.

IEC. (2014a). *Railway applications — Urban Guided Transport Management and Command/Control Systems — Part 1: System Principles and Fundamental Concepts*, IEC 62290-1:2014. International Electrotechnical Commission. Geneva.

IEC. (2014b). *Railway applications — Urban guided transport management and command/control systems — Part 2: Functional requirements specification*, IEC 62290-2:2014. International Electrotechnical Commission. Geneva.

Lucic I. (2015). *Risk and Safety in Engineering Processes*, Cambridge Scholars Publishing ISBN (13): 978-1-4438-7077-1.

RailSystem. (2022). *Communications-Based Train Control (CBTC)*, https://railsystem.net/communications-based-train-control-cbtc/, Accessed 4th July 2023.

SEBoK. (2022). *Guide to the Systems Engineering Body of Knowledge*, v. 2.7, released 31 October 2022, https://sebokwiki.org/wiki/Operational_Scenario_(glossary). Accessed 4th July 2023.

Sparx Systems. (2022). *UML 2 Tutorial – Sequence Diagram*, https://sparxsystems.com/resources/tutorials/uml2/sequence-diagram.html. Accessed 4th July 2023.

Appendix A. EUC Hazard Descriptions

Table 8 ~ EUC Hazard Descriptions

ID.	Pre-existing Hazard	Description
Hp#1	Conflict between any pair of train trajectories (see note below table)	This hazard is about the separation between trains. As a state of the railway, it exists whenever the intended movement (e.g. planned missions / perturbed running) of any two trains would result in the trains being at the same location at the same time, i.e. a **collision** would result if nothing at all were done to prevent it.
Hp#2	Conflict between a train's trajectory and track configuration	This hazard is about the relationship between the intended routing of a train and the configuration of the track elements. It exists whenever the intended movement of any train would result in the train passing through an incorrectly-configured set of points or level-crossing lights / barriers — **derailment** and/or **collision** could result if nothing at all were done to prevent incorrect route setting.
Hp#3	Train speed exceeding capabilities of the track infrastructure and/or train	This hazard is about the relationship between the speed of a train and the ability of the track elements to support it. It exists whenever the speed of any train exceeds the capability of the track, taking account of the permanent, intrinsic (e.g. curves and junctions) or temporarily-degraded characteristics of the track (e.g. buckled or broken rail) **derailment** (or **collision**, in the case of over-speed at the end of track) could result if nothing at all were done to prevent over-speeding of the train.
Hp#4	High and/or uneven acceleration / deceleration of a train	This hazard relates the effect on passengers on trains due to sudden train movement. It exists during lurching, jerking, or sudden rapid deceleration, which could result in **passenger falls** with the possibility of injury, serious injury or (exceptionally) death.
Hp#5	Conflict between train profile and fixed structure, *except* as the result of excessive train speed (Hp#3) or damage to structure (Hp#10)	This hazard covers the relationship between the intended route and speed of a train and structure gauge. It exists whenever it would be possible to route a train through, or past, a fixed structure whose gauge is incompatible with the kinetic envelope of that train (as determined by, *inter alia*, its size / shape and speed), and would result in a **collision** if nothing at all were done to prevent it. It excludes potential collisions with fixed structures arising from derailment (see other, derailment-related hazards), excessive speed of the train, and collisions arising from failure of fixed structures (see Hp#10)

ID.	Pre-existing Hazard	Description
Hp#6	Conflict between a train's trajectory and non-fixed obstacles or unauthorised persons on track	This hazard concerns the *unexpected* presence of objects, large animals or unauthorised persons on the running railway such that they could make contact with a passing train. Depending on the physical properties of object concerned, it could lead to: derailment; damage to the leading cab, with the possibility of train-operator injury, serious injury or even death; or serious injury / death to persons on the track.
Hp#7	Conflict between a train's trajectory and workforce personnel / vehicles on track	This hazard concerns the *planned* presence of workforce personnel or vehicles / equipment on the running railway such that they could make contact with a passing train if nothing at all were done to prevent it.
Hp#8	Passengers attempt to exit a train outside a station	This hazard covers the possibility of passengers falling out of a train due to: - train doors being opened too early on entry to a station; - a train departing with a door or doors open; - train doors being opened outside of a station; or - carriage separation
Hp#9	Passenger embarkation / disembarkation at platform	This covers possible incidents associated with normal entering or alighting from trains at a station. It includes: - train doors being opened on the side away from the platform leading to passengers getting off the train on the wrong side or falling out of the train on to the track; - train doors which are on the same side of the train as the platform, but which are not adjacent to the platform (i.e. the train is longer than the platform, or is not correctly berthed) being opened and passengers falling out of the train; - train doors opening at a closed station except where done deliberately (e.g. to evacuate passengers from platform or train); - a passenger being hit by closing door; - a passenger (or passenger's clothing) being caught in door of a stationary train, which then moves off, dragging the person along the platform; - slips, trips and falls associated with the gap between the train and the platform.

ID.	Pre-existing Hazard	Description
Hp#10	Structural failure of track elements, tunnels, bridges, etc.	This hazard addresses the threat of **collision** to trains (and its Passengers and on-train Workforce) from failure of structures, including: - unsound track elements; - unsound / unsecured tunnel; - unsound / unsecured under-bridge / culvert; - unsound / unsecured over-bridge. It excludes the direct effects of such failures on members of the public. It also excludes failure of other railway structural assets (e.g. signalling or electrical structures), fallen trees, etc., all of which are covered by Hp#6.
Hp#11	Personnel exposure to potentially lethal voltage	This hazard addresses the threat to people of **contact with lethal voltages** from electrical power supplies.
Hp#12	Passengers too close to, or fall/jump off, platform edge	This hazard concerns the possibility of passengers at a platform being struck or run over by a train due to passengers: - standing too close to the platform edge or otherwise infringing the kinematic envelope of the train; - falling off (or jumping of) platforms; - crossing the lines at a station (where unauthorised only).

Note: Conceptually, a train's "trajectory" is the path and speed profile that the train *intends* to follow at any point in time, in the absence of any instructions to the contrary.

Appendix B. Functional Safety Requirements - Examples

As an example, the following table lists all FSRs for the Safety Related Systems that have been derived for OSF#1, "*Establish & Protect a Safe Route for each Train Movement*", in the analysis at Sub-section 4.6, and shows traceability back to the related OSRs set out in Table 4.

The requirements themselves have been adapted from the UGTMS standard (IEC 2014b), and represent a full list for each Safety Function, but under normal operating conditions only.

The traceability shown is to the Overall Safety Requirements (OSRs) related to OSF#1.

Table 9 ~ FSRs for SRSs for OSF#1 -
Establish & Protect a Safe Route for each Train Movement

ID	Safety Requirement	Traceability
SRS#1	**Set & Protect Route Elements**	OSF#1
SF#1.1	***Reserve, Set & Lock a Standard Route***	
FSR1.1.1	For the route to be reserved, ATC shall reserve all the route elements required based on the route origin and route destination, including elements required for flank protection, and for overlap.	OSRs 1.1 & 1.2
FSR1.1.2	The reserved status of a route element shall be provided by ATC to other functions and Service Control Centre.	OSRs 1.1 & 1.2
FSR1.1.3	ATC shall move a reserved movable route element to the desired position if it is not already in that position, not occupied by a train and not blocked against moving.	OSRs 1.1 & 1.2
FSR1.1.4	If a movable route element does not reach the desired position in a predefined time, ATC shall initiate a failure message to this effect.	OSRs 1.1 & 1.2
FSR1.1.5	ATC shall lock all route elements in a route to be set if they are confirmed in the required position.	OSRs 1.1 & 1.2
FSR1.1.6	ATC shall not set a route which would allow a train to enter a route for which it is not suited.	OSR1.4

ID	Safety Requirement	Traceability
SF#1.2	***Supervise Route***	
FSR1.2.1	ATC shall monitor the status of all route elements to confirm that they are in the required position and locked.	OSRs 1.1 & 1.2
FSR1.2.2	ATC shall provide the status of each route. to other functions and Service Control Centre.	OSRs 1.1 & 1.2
FSR1.2.3	The entrance to a route shall be prohibited by ATC in response to a safety related manual input.	OSRs 1.1 & 1.2
SF#1.3	***Maintain Route Locking***	
FSR1.3.1	ATC shall determine a train approach area in front of a route origin for which a Movement Authority has been given. The approach area shall cover an area which is longer than the operational braking distance, allowing for any human or system reaction time.	OSRs 1.1 & 1.2
FSR1.3.2	ATC shall ensure that the status "route locked by approach" prevents the immediate release of the route: • if a train is in the approach area and a movement authority has been given to the train, *or* • if a train has entered the route (with or without movement authority).	OSR 1.3
FSR1.3.3	ATC shall ensure that moveable route elements (e.g. points, etc.) that are occupied by trains are prevented from moving, regardless of whether or not the route is set.	OSR 1.3
FSR1.3.4	ATC shall ensure that the route elements in front of a train are maintained locked as soon as the train has entered the set route.	OSR 1.3
FSR1.3.5	ATC shall ensure that moveable route elements that are "blocked against switching "remain in that state until released by manual input related to the need to block the elements in the first place.	OSR 1.3
FSR1.3.6	ATC shall ensure that moveable route elements in a Recovery Route remain locked until the route has been removed.	OSR 1.3
FSR1.3.7	A moveable route element that has been locked by route-setting or manual input shall not be released until all routes / manual inputs that caused the element to be locked in the first place have themselves been released / removed.	OSR 1.3
FSR1.3.8	ATC shall not release route elements that are providing flank protection for a route until the route itself is released.	OSR 1.3
SF#1.4	***Release Route***	
FSR1.4.1	A route may be released **only** if and when **all** of the conditions for maintaining it locked no longer apply.	OSR 1.3

ID	Safety Requirement	Traceability
SRS#4	**<u>Authorise Train Movement</u>**	OSF#1
SF#4.1	***Determine Limit of Movement Authority***	
FSR4.1.1	ATC shall determine for each train the limit of its movement authority (LMA) based on the most restrictive of the following potential conflict points: • Limit of safe route, • Limit based on safe train separation, • Limit based on the physical infrastructure (e.g. end of track), • Zones of protection.	OSR1.5
FSR4.1.2	In the event of a loss of safe route once a movement authority has been issued, ATC shall pull back the LMA to the new limit of safe route.	OSR1.5
SF#4.2	***Determine Train Protection Profile***	
FSR4.2.1	ATC shall determine a Train Protection Profile for each train, to prevent it from overrunning its LMA, or exceeding the applicable speed limits within its LMA.	OSR1.5
FSR4.2.1	The Train Protection Profile shall be determined by the applicable Safe Braking Model — an analytical representation of a train's performance while decelerating to a complete stop, allowing for a combination of worst-case influencing factors (gradient & adhesion, etc.) and failure scenarios.	OSR1.5
FSR4.2.2	The Safe Braking Model shall ensure that an ATC equipped train will always stop within a distance not greater than that guaranteed by the Model.	OSR1.5
FSR4.2.3	ATC shall calculate the train-protection profile that results from the most restrictive of all safety-related constraints applied to the ATC-equipped train.	OSR1.5
FSR4.2.4	ATC shall enforce speed limits for the whole length of the train.	[n/a]
SF#4.3	***Authorise Movement of Reporting Trains***	
FSR4.3.1	If a Train Protection Profile with permitted speed greater than zero is established, train movement shall be allowed, up to next LMA.	OSR1.5
FSR4.3.2	Each train movement authorised by ATC shall be within the constraints of the applicable Train Protection Profile.	OSR1.5

ID	Safety Requirement	Traceability
SRS#5	**<u>Supervise Train Movement</u>**	OSF#1
SF#5.1	***Determine actual train speed***	
FSR5.1.1	ATC shall detect and determine the actual train speed, taking into account the effects of speed-measurement inaccuracies.	OSR1.5
FSR5.1.2	ATC shall determine the zero-speed status within the predefined tolerances of the speed measurement system.	OSR4.1
SF#5.2	***Supervise Safe Train Speed***	
FSR5.2.1	ATC shall supervise the actual speed of trains to ensure that each train remains within its Train Protection Profile.	OSR1.5
FSR5.2.2	ATC shall trigger Service braking in accordance with the warning profile in order to respect the Train Protection Profile and to avoid Emergency-brake intervention.	[n/a]
FSR5.2.3	ATC shall automatically release the Service brake during deceleration if actual determined train speed returns below the warning profile.	[n/a]
FSR5.2.4	If the determined actual train speed is higher than the speed permitted by the Train Protection Profile, ATC shall trigger emergency braking.	OSR1.5
FSR5.2.5	ATC shall provide two possibilities for automatic emergency brake release: • if, during deceleration, actual determined train speed returns below the train-protection profile provided there are no other conditions for triggering the emergency brake. • if actual train speed is determined as zero and there is no other triggering condition,	[n/a]
SF#5.3	***Supervise Safe Train Direction***	
FSR5.3.1	ATC shall detect an unauthorized movement of the train in case of travel of the train against the authorized direction of travel beyond a predefined distance,	[n/a]
FSR5.3.2	When unauthorized movement of the train against the authorized direction rollaway is detected, ATC shall apply the emergency brake,	[n/a]
FSR5.3.3	In the event that a *moving* train receives a Movement Authority that is contradictory to its direction of travel (i.e. is "behind" the train), the train shall: • emergency brake to a standstill • not accept the Movement Authority • report its inability to accept the Movement Authority, to Service Control.	[n/a]

ID	Safety Requirement	Traceability
SF#5.4	***Supervise Movement-Authority Validity***	
FSR5.4.1	In case a movement authority accepted by the train exceeds its validity period (e.g. due to data communication failure), ATC shall either: • pull back the movement authority limit to the first conflict point ahead of the train, or • stop the train immediately.	OSR1.5
FSR5.4.2	In the event that the train's Movement Authority is cancelled, the train shall emergency brake to a standstill.	[n/a]
SF#5.5	***Overrun Protection***	
FSR5.5.1	ATC shall supervise the actual position of each ATC-equipped train against its LMA and initiate an emergency braking in the event that the LMA is exceeded.	OSR1.5
FSR5.5.2	ATC shall restrict the movement authority of ATC trains that are in conflict with an unauthorised movement of any train when such an unauthorised movement is detected.	[n/a]

Appendix C. Operational Scenarios — Examples

C.1 Operational Scenarios for *Normal* Operations

Table 10 provides examples of Operational Scenarios for *normal* operations conditions, as discussed in Sub-section 4.6.4.

Table 10 ~ Example Operational Scenarios for *Normal* Operations

SC No	Description	Related Actors	Trigger In:	Trigger out:	Related States
01	Empty ATC passenger train ready to enter service from depot / siding (train berthed outside ATC signalling area	SCC, ATC Train, SRS#1, SRS#2, SRS#4, SRS#9	SCC requests Train power-up	Train registered on ATC system and position ready to exit depot / siding for first Mission of day	ST001, ST002, ST003, ST004
02	AC Train undertaking a System-protected Movement, between station stops (following a non-reporting train)	ATC Train, non-ATC Train, Non-ATC wayside SRS#1, SRS#2, SRS#4, SRS#9	Train leaves previous station limits (SC-03)	ATC Train approaches next station stop (SC-03)	ST005
03	ATC Train makes scheduled station stop.	ATC Train, SRS#1, SRS#2, SRS#4, SRS#9	ATC Train approaches next station stop (SC-02)	Train clear of station limits (SC-02)	ST006
04	Train repositions to reversing location, for next mission.	ATC Train, SRS#1, SRS#2, SRS#4, SRS#9	Train leaves final station limits (SC-03)	Doors closed. Train ready to depart reversing location (SC-02)	ST007, ST003
05	Train repositions to depot, after completing final mission of day.	ATC Train, SRS#1, SRS#2, SRS#4, SRS#9	Train leaves final station limits (SC-03)	Train powered down in depot siding	ST007, ST002, ST001

SC No	Description	Related Actors	Trigger In:	Trigger out:	Related States
06	Route setting and junction management in ATC areas	SCC, ATC Train, non-ATC Train, Non-ATC wayside SRS#1, SRS#2, SRS#4, SRS#9	Timetable implementation	Rear of train's Virtual Occupancy clears junction	ST005
07	Train exits depot to join mainline on first mission	ATC Train, SRS#1, SRS#2, SRS#4, SRS#9	Train registered on ATC system and position ready to exit depot / siding for first Mission of day	Train running according to timetable	ST004, ST005

C.2 Operational Scenarios for *Abnormal* Operating Conditions

Table 11 provides examples of Operational Scenarios for *abnormal* operating conditions, as defined in Sub-section 4.6.5.

Table 11 ~ Example Operational Scenarios for *Abnormal* Operating Conditions

SC No	Description	Related Actors	Trigger In:	Trigger Out:	Related States
016	ATC Train performs recovery of failed train.	SCC, ATC Train, SRS#1, SRS#2, SRS#4, SRS#9	Trains coupled, registered in ATC system and ready to move	Train reaches recovery location	ST011
033	Service Control applies Temporary Speed Restriction.	SCC, ATC Train, non-ATC Train, Non-ATC wayside SRS#1, SRS#2, SRS#4, SRS#9	TSR initiated by SCC	TSR in force – steady state	ST009
079	Detrainment of passengers, train not berthed in platform (taking passengers off train on foot)	SCC, ATC Train, SRS#1, SRS#2, SRS#4, SRS#9	Event necessitating detrainment	Passengers evacuated to safe location	ST010

SC No	Description	Related Actors	Trigger In:	Trigger Out:	Related States
096	Movement of non-communicating ATC trains	SCC, ATC Train, non-ATC Train, Non-ATC wayside SRS#1, SRS#2, SRS#4, SRS#9	Non-communicating ATC train needs to move	Non-communicating ATC train completes move	ST011
100	Service Control initiates unplanned station stop	SCC, ATC Train, SRS#1, SRS#2, SRS#4, SRS#9	SCC needs to stop a train at a non-timetabled station / platform	Train berthed at platform; doors ready to open if required	ST009, ST010
108	Passenger evacuation from train fire	SCC, ATC Train, SRS#1, SRS#2, SRS#4, SRS#9	Fire on train	Passengers rescued	ST010
109	Train / passenger rescue from wayside fire	SCC, ATC Train, non-ATC Train, Non-ATC wayside SRS#1, SRS#2, SRS#4, SRS#9	Fire on wayside	Passengers rescued; train recovered	ST010

This collation page left blank intentionally.

ISSN 2754-1118 (Online)
ISSN 2753-6599 (Print)

Chasing the Black Swan

Malcolm Jones

Atomic Weapons Establishment (AWE), Aldermaston, Berkshire, UK

Abstract

The term Black Swan is a familiar concept in the context of high consequence operations. There is the continual concern that there may be an 'as yet' undiscovered flaw or lack of understanding in the design of a product, process or facility that could lead to a catastrophic event. Concern lies in the potential incompleteness of understanding of any design concept, implementation and associated assessment. Given that 'absolute confidence' may never be possible, the question arises as to how best to continue to search for such a possible flaw with a view to subsequent removal or mitigation. This at first sight appears to be a process without end but the level of commitment must be balanced against the detrimental consequence that could ensue given that a Black Swan might exist. But when is 'enough-enough'? This subject is covered in the context of the ownership of nuclear warheads where the Black Swan can indeed be catastrophic should it exist. The paper is framed somewhat in terms of what can be learned from the general literature associated with Black Swan thinking.

1 Introduction

Aspects of this subject have been covered in previous papers (Jones 2016) (Jones 2017), which have taken different perspectives of an independent strength in depth approach to safety and its assessment. The first paper concentrated on organisational structures in relation to the necessary organisational levels, each having independent responsibilities for safety ensurance and assurance[1], coupled with final decision-making responsibilities and of course each having the appropriate level of technical capability and experience. The overall organisational responsibility is that of ensuring a safe product, process or facility with appropriate technical and evidential support, with the level of scrutiny proportional to the potential consequence of getting it wrong.

The first organisational layer is responsible for ensurance through making the safety case with a supporting evidence base. The second independent structure scrutinises and challenges this case for appropriate depth and completeness, including assuring that the appropriate level of expertise has been applied and that the evidence and analysis offered is complete and not flawed.

The third independent organisational layer is responsible for final decision making, having taken into account and scrutinising the ensurance and assurance evidence provided by the

[1] We use the term "ensurance" to mean the demonstration that a design, etc., has met its requirements; whereas "assurance" is the independent challenge of whether the ensurance is valid.

© Crown copyright (2023), AWE. This material is licensed under the terms of the Open Government Licence except where otherwise stated. To view this licence, visit http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3 or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gov.uk.

first two layers and in turn, adding its own independent assessment based on a complementary fund of knowledge and experience. This third layer will also manage and resolve any disagreement arising from the views arising from the first two layers. Some of the fundamental competences needed for correct operation of this structure were identified in Jones (2016) together with some of the remaining difficulties.

However, our historical experience suggests this overall approach alone may not be sufficient for ensuring the absence of a Black Swan for a product such as a nuclear warhead. For this reason, a fourth layer is introduced, see Figure 1 (Jones 2016). That is a somewhat undefined layer which continues to probe into the product, process or facility with the intention of taking a 'what if' and 'expecting the unexpected' mind-set. This layer's job in essence is never finished and sets its target in the broadest sense. More is said later.

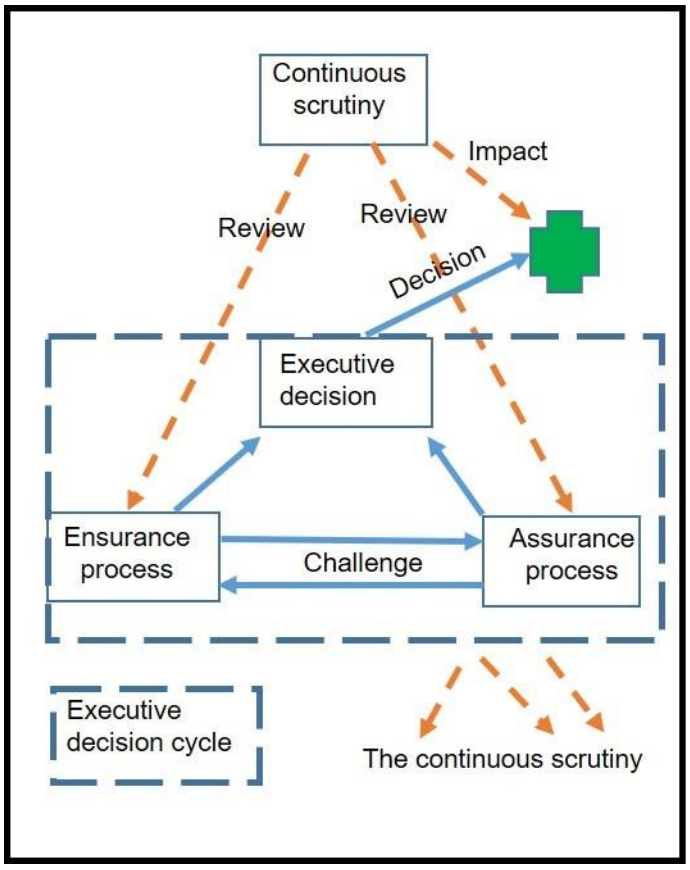

Figure 1 ~ The Continuous Assessment Process

Note that the Green Cross in the diagram is the potentially revised executive decision given what the fourth layer has unearthed as a result of continued scientific investigations; "Impact" is the impact on the executive decision given new information unearthed by the fourth layer.

The second paper covered two independent technical strategies for safety ensurance and assessment (Jones 2017). The first was based on an independent strength in depth technical analogue of the organisational structure. This is based on a deterministic approach founded on fundamental and independent levels of technical defence which is intended to include a hedge against uncertainty. The second approach complements this with a realistic, but conservative, evidential based numerical risk assessment looking for compliance with risk standards. Of course, any risk assessment will contain an element of foundational judgment for supporting such an analysis. In principle, this judgement could

be in error and for this reason an acceptable level of conservatism has to be included in order to cater for any remnant uncertainty. This risk assessment relates directly to a particular (or set of) design solution as opposed to the deterministic approach which is based simply on 'how to design for safety'. The latter risk approach aims more for a demonstration of meeting acceptable risk standards given the design solution.

The two approaches have elements of fundamental difference but in concert compensate for individual potential limitations and statutory requirements. From a philosophical point of view the deterministic approach does offer the better of the two approaches for protection against the potential of a Black Swan because of its greater potential for protecting against any limits in our 'complete knowledge'. The risk assessment is necessary because of the societal demand to understand 'what the level of risk is' and whether it is tolerable in meeting accepted standards. However, the risk assessment inevitably bases its approach on an 'accepted level' of understanding with some conservatism included to cover any potential deficiency in knowledge.

There is some value in the argument that, given a history of a product or process performing satisfactorily over a given period, this adds further to the confidence that a Black Swan may well not exist. However, the value of such evidence must be set against the level of potential detrimental consequence. For example, if the requirement is that a major consequence should not have a frequency of greater than a 1 in 100 years, then a ten-year successful history only gives additional confidence that all seems well. In fact, if the frequency requirement for a catastrophic event is far more demanding, then typical design history perspectives may be somewhat less comforting. History also shows us that past success does not always support future success.

2 Historical Experience

History shows us that many, if not most, catastrophic events occur through an 'unfortunate' sequence of linked events rather than from a single technical cause. Such previously undiscovered sequences are often related to an incomplete knowledge of the range of possible states leading to the propagation of the failure sequence or, perhaps more correctly, the level of dependency between such states in such a sequence. For example, an initial technical failure will not propagate if all states potentially influenced by such a failure have been fully identified, characterised and configured through independence to terminate such a sequence. The problem arises when there is incomplete knowledge about the range of states that may be mutually or dependently influenced. So incomplete information is the enemy of safety assurance of termination of a mishap occurrence. This is often pictorially displayed in the Swiss Cheese model. In the general case the states involved may take on a wide range of types.

For example, in a simple technical product an initial analysis may appear relatively simple, i.e. that of tracking the well-known and fully-defined influenced states together with their known mitigating/independent resilience. However, the response of these states may in turn be influenced by external actors, for example environments at the time of the initial flaw and without this information the response robustness of the argument for terminating the sequence prior to the mishap will be questionable. Of course, nuclear warheads need to remain safe given a wide range of potentially severe environments. To treat any real case, it is necessary to establish, as far as possible, a complete definition of the 'system' of all states and influences. This is perhaps analogous to the goal of fundamental physics of seeking a complete understanding of all of nature's subatomic particles and forces and their potential interactions. As the 'system and influences grow' so does the difficulty of

being able to fully identify and characterise the response of such states given any initiating event. In essence, complexity can become the enemy of safety and enhances the difficulty of identifying a Black Swan should it indeed exist. This complexity is perhaps best illustrated in simple form by a fictitious, but not improbable, example given in the form of a somewhat dated cinema film, "Fate is the Hunter" (Nelson 1964), shown during the 35th International System Safety Conference (2017) by John Rankin, who used it by way of example. It is paraphrased as follows:

Following an aircraft take-off one of the engines was struck by a bird which was rarely found in that part of the world — an unlikely initiator. The plane suffered a jolt due to the loss of that engine and as a result a cup of coffee, just supplied to the Captain, spilled onto an equipment box. Because the box was unsealed, coffee leaked onto the underlying equipment. The coffee was somewhat conducting in nature and as a result this interfered with underlying critical circuits, which led to the loss of the other engine. Because of aircraft congestion at the airfield at that time, the Captain took the reasonable option of landing on a nearby suitable beach. It so happened, unknown to the captain, that a pier on that beach, which had been scheduled for removal the previous week, was still in place, because of a decision by the contractor that there was no real need to hurry. Hence the fateful collision.

Although the initial verdict was judged to be human failure, careful analysis showed this not to be the case. The sequence is fictitious, but it does exemplify the need to fully understand what the 'system and the influences' really consists of and hence all of the potential interacting events given a somewhat unexpected fault initiator. Who would initially have identified the potential presence of a cup of coffee, the importance of an unsealed equipment box, congested airport conditions at the time and the continued existence of the pier that should have been absent?

Simplicity in the 'bounding of the system' is an important goal for minimising the opportunity for failing to fully characterise it. This in fact represents an important element in the safety approach to nuclear warheads: bounding the system (in the safety sense) as far as possible and aiming for simplicity, independence and clarity in the safety argument. Some of these aspects are covered in more detail later in this paper. The additional lesson learned from this fictitious example is that the sequence initiator was an abnormal environment, i.e. is an 'insult' to the technical system rather than a design flaw, although one could argue that design flaws helped the sequence to propagate. In fact, history shows that many sequences which have led to mishaps come from initial 'insult' initiators. This lesson is not lost on the strategy for minimising the opportunity of a Black Swan manifesting itself in nuclear warheads. The strategy is based on both the effort expended in both preventing such insults and the safety resilience given such an insult.

3 General Background to the Black Swan Term

3.1 Origin of the Phrase 'Black Swan'

A rare event, based on the belief widely held in England in the 1600s that swans could only be white. All swans in England at that time were white. The phrase "a Black Swan" was a metaphor for "that which could not exist". In the late 1600s, in Australia, black swans were discovered.

3.2 Current Understanding of the Phrase

An event judged through best established inductive/inferred logic to be assessed as rare — but this process does not remove the possibility of occurrence.

1. That the knowledge that we base our inductive assessment on is not quantitatively or logically correct. Its impact on risk assessment.
2. Or, more importantly that the scope of our inductive process is not complete, i.e. there are some possibilities we have not yet visualised. *Its impact on both risk assessment and defence in depth*.

The bottom line is, given the application of our best knowledge, a quantitative assessment may be broadly realistic but does not discount the possibility of an occurrence. In most cases this quantitative assessment is sufficient and the whole subject areas of 'cost benefit and insurance risks' work on this principle. However, this approach may be inappropriate if the Black Swan event has a major or catastrophic consequence. In this case there are two emerging issues:

1. The incompleteness in the quantification of the probability of mishap occurrence may be small but this delta does matter.
2. That the true nature and impact of the mishap may itself not be understood and underestimated. Although there may be a reasonable description of the mishap, the real follow-on consequences are usually less well understood, and often turn out to be of a far greater nature than first anticipated.

Hence, the context of the Black Swan is that it applies to a mishap which is assessed to be somewhat improbable, but cannot be discounted, and that its outcome can be severe or catastrophic and not fully understood. In fact, in a general sense, it might be unclear as to what the mishap itself may be and this is typified by the lessons from economic/financial traumas.

3.3 Some Terms and Definitions

Unexpected — The 'best analysis', suggests that the event has a low probability of occurrence. Such analysis cannot claim to be absolutely free of incompleteness and as such cannot assure us it cannot happen. There is often a quantitative definition of what is meant by unexpected.

Inconceivable — One cannot imagine how it can possibly happen, or believes that there is some fundamental reason why it cannot happen. This in principle is a human construct, and does not completely remove the possibility of a high consequence event.

Consequence — This relates to the range of impacts (detrimental in this case) given the event and in the case of safety this will take many forms:

- Death or harm
- Environmental damage
- Loss of asset
- Financial loss
- Reputational loss
- Political harm
- *Et cetera*.

Black Swans are generically spoken of in the context of a major detriment and, for nuclear warheads, can be associated with a major issue in relation to performance or safety. In the latter case this can be directly related to the occurrence of the worst-case catastrophic event itself, or to realisation that there may be a major flaw in the safety argument which requires urgent mitigating action. Of course, this is a key issue in the continued ownership and safety scrutiny of a stockpile of nuclear warheads.

4 Some Differing Categories of Black Swan and AWE's Culture

4.1 Known — But Not To The Degree Necessary

Knowledge about such a possibility is available but is not sufficient for correct judgement on probability and/or consequence. Of course, there is always the judgmental problem of when 'enough is enough' in relation to the depth of scrutiny which is necessary for elimination of the Black Swan's possibility.

With regards to detrimental event occurrence: The warhead programme has strived to avoid falling into this category. It has taken a very much 'what if' and 'expecting the unexpected' approach. All known possibilities are subject to in-depth scrutiny. If there is any reason why an issue has not been completely closed out, it is subject to continued scrutiny and of course this may give rise to a necessary change of design. We continue to strive to establish a more comprehensive understanding of the underlying characteristic of the known issues, whether they be of technical, human factors, implementation or governance nature. We retain a culture of assuming that we may not have reached the end of the road towards this goal and there is always more to be done — a culture of continued in depth assessment[2].

4.2 Unknowns — But Knowable Unknowns That Could Have Been Looked For Or Known By Others

The problem here is that such unknowns may have been regarded (erroneously) as being of no consequence, e.g. for safety.

With regards to detrimental event occurrence: The relevance to warhead activity follows from the comments under the last heading in the sense that there should be no assumption, without sufficient evidence, that the knowable unknowns can be discounted. If the unknowns are knowable then this gives us a direction in which to focus further effort and scrutiny with the goal of reaching a position where such unknowns are converted into 'fully characterized', and their impact determined. Included in this is the need to be aware of what's happening elsewhere in the world where similar technologies may be exercised, and where there are technology enhancements taking place which can have an advantage in enhancing our safety. Of course, in this respect one covers the whole associated subject area included in STEM (Science, Technology, Engineering, and Mathematics) where it is necessary to keep a close eye on worldwide developments in both industry and academia, which can be incorporated to our advantage. Our close collaboration with our US colleagues plays a major mutual role in this.

[2] Note that the fourth layer is not specifically responsible for assessing safety, but has a more open remit for probing further and further into the scientific and technical aspects that underpin nuclear science and, if something is yet to be discovered that could be a 'Black Swan', it is here that it will be found.

In addition, one needs to keep abreast of 'state of the art' developments in the range of methods for safety assessment and assurance, much of which are bound up in the general subject area of System Safety and as such where we need to be expert practitioners. Of course this again is a constantly moving and evolving activity with an acknowledgement that there will be no perfection. Any 'judgement' that we have reached a position of having identified and characterized such known unknowns is subject to constant challenge. Again the 'what if' attitude and 'expecting the unexpected' culture at AWE does provide a strong defence against this form of failure. Of course, the primary element of protection relating to 'known by others' comes through collaboration with our US colleagues and other work, particularly in the nuclear industry at large. The aim here is to establish confidence that no potential influencing subject area is left untouched.

4.3 Unknown — But Unknowable Unknowns

No absolute basis for being able to make any assured prediction for occurrence expectation or consequence — *a true Black Swan!* Nevertheless, even here 'confident' predictions may well be made but, of course, with the potential for failure due to ignorance.

With regards to detrimental event occurrence: This presents the most difficult aspect with regard to warhead safety assurance in that, if such flaws exist in our knowledge, we have little or no guidance of where to look further. The best practical protection against this uncertainty comes from the independent strength in depth approach to design, ownership and assessment both in terms of the technical and organisational aspects[3]. The best overall solution we are able to come up with in this respect, given no clear guidance of where further to look, is a continued commitment to ensure that we have a cadre of high calibre staff working at the cutting edge of all of the relevant technologies, processes and assessment methodologies and giving them sufficient freedom and encouragement 'to probe' into 'looking for the unexpected'. This activity is not solely restricted to 'in house' activity but also includes keeping a close look at developments of interest in the external world and how best to take advantage of such developments. Such unknowns are not likely to be uncovered through standard process and routine. From a business perspective, this may appear to be an untargeted, unproductive and non-cost-effective overhead — but we fail here at our peril. This Continuous Scrutiny process is identified in Figure 1. *AWE's culture strongly supports and encourages this approach.*

5 5 Lessons from the Open Literature

5.1 Nassim Nicholas Taleb

Taleb is well known for his work in this area and describes a Black Swan (Taleb 2008) in terms of:

1. It is unpredictable and as such cannot be truly sentenced as low probability;
2. It causes an extreme or catastrophic impact; and
3. It is retrospectively predictable — with the warning that there can be a tendency to simplify (possibly erroneously) the post event analysis of the cause. *There may be a telling lesson here for us in relation to the Review, Learn and Improve*

[3] Beware of 'groupthink' in which people strive for consensus; the claim of independence is always subject to credible challenge

> *actions we undertake — we need to be clear that we really do have the correct post event analysis — otherwise it can happen again.*

Taleb's interest in unpredictability comes from his association with economics and the financial industries where the activity of prediction has been shown on many occasions to be unwarranted with major detrimental consequences. There may be some success in predicting the near future, but this gives no assured guidance for the longer term and the longer term may not be too extensive in this context. This aspect is somewhat like weather prediction where longer term prediction is also fraught with uncertainty. This can manifest itself via so-called Chaos Theory. In both cases predicting the future is bedevilled by the vast number of interacting contributors of, not always perfectly defined, knowledge and starting conditions. In the case of the economic and financial industries the situation is further complicated by the emotional response of people, whereas weather prediction is primarily governed by the laws of physics. *It is salutary to note that a nuclear weapon (or even a warhead) design, including its processing and ownership, is itself a somewhat complex subject area with inherent complicated relationships. However, a key lesson from Taleb's experience is that we should strive to limit complexity*[4] *(in respect to its relationship to safety) because of its impact on the ability to predict with confidence. As a result, we continue to strive to limit complexity and to gain as complete an understanding as possible of what we have — with the mind set of 'what if' — is it complete?*

There have been many attempts to model economic and financial futures, some based on a statistical basis, for example on the normal distribution for occurrence probability and range of consequence. In the context of financial industries, the 'well known' distribution peaks can be associated with the 'currently understood' in the relatively short term and the tails, with all their uncertainties, can be represented by the longer term 'judgements'. Because of the complex nature of the financial industries, these 'tails' often turn out to be completely misleading. Distributed approaches are sometimes applied to warhead safety assessments. Such distributions usually have far more evidential support near the peaks and less in the tails, where the low probability high consequence predictions are made and where prediction can be in significant error. This is where the Black Swans, if they exist, may lie. *The safety of warheads depends on the evidential confidence in the application of such distributions, where such distributions need to be applied.*

5.2 Taleb's View in Combatting Uncertainty

Because of the complexity associated with the financial industry, Taleb suggested that the effort expended in trying to prevent catastrophic occurrence was not well spent. He felt that catastrophes were inevitable and most effort should be best directed towards robustness in coping with Black Swan events when they do occur, rather than on predicting or preventing their occurrence. His experience was associated with economic and financial subject areas where his assumption (based on substantial historical evidence) was that these events will occur, and one should have plans in place on how to deal with them and mitigate the consequences. Given the assumption that the event will occur, Taleb's approach appears correct. *However, the priority for nuclear warheads lies firmly in the direction of preventing the occurrence and particularly in relation to the 'worst-case' occurrence of Inadvertent Nuclear Yield, rather than prioritising on the mitigation actions following such an event, if indeed this was realistically possible. For this reason, we concentrate on limiting complexity, particularly in the context of the principal safety*

[4] However measured

arguments, in order to enhance transparency and confidence in the evidence to support an assessment of a very low occurrence probability.

5.3 Charles Perrow

Perrow is well known for his work in accident theory (Perrow 2007). He provides another guide as to how Black Swans can arise, and what approaches should be put in place to avoid their occurrence. He has spent a great deal of his career in the field of historical accident assessment and the reasoning behind their occurrence. These have covered a wide range of subject areas, but their applicability are generally pertinent to nuclear warhead safety. The principles that clearly emerge from his work are again simplicity, clarity, and independence. These are evidenced by his observations and recommendations noted below.

1. *Complex, interconnected, and highly-coupled systems should be expected to fail*

Response in warhead design: Warheads by their very nature have to be interconnected but this is different in the sense of being highly coupled in the context of safety. For example, interconnectivity in the sense of arming safety requires a unique set of authorizing actions/conditions for 'fault progression' and their absence prevents unintended fault progression. This principle also applies to other aspects of the warhead. As such in the safety sense there is limited inadvertent connectivity. The same is true for the AWE safety organizational structure in that, although there are interconnections, they are designed to be independent in the safety ensurance and assurance sense. Design also aims to minimize complexity, and this helps towards unintended (undetected) safety interconnectivity. For example, safety systems strive to minimize component counts (and particularly those that are safety-related). In addition, principal safety arguments are based on a demanding requirement to demonstrate mutual independence between safety systems including the fundamentals of principles and implementation. Coupled with this is a strategy aimed at making the principal safety argument transparent and simple, as opposed to being complex in nature. The more complex the safety argument the more challenges that can be realistically raised, and the greater the difficulty in providing the appropriate level of assurance. Both simplicity and independence minimize the opportunity for failure through unintended coupled and undetected paths. In turn simplicity and independence ease the burden of providing safety assurance given that one is always tasked to continue to look for all the 'possibilities' leading to mishap. It is certainly not a stance of taking more and more comfort from what we have achieved so far. A strong continuous probing, analysis, testing and surveillance culture is key in relation to early detection of any flaws that might exist or may be developing over time, and is supported with a commitment to maintaining a strong cadre of experts probing in the relevant technology areas.

2. *Focus on those aspects with catastrophic consequences*

Response in warhead design: Although we focus on safety across the board, we nevertheless concentrate particularly on the worst-case events of Inadvertent Nuclear Yield and major radioactive material release, and consequently go to great lengths to prevent their occurrence both by way of design, associated processes and independent organizational assessment activity. Again, the prevention of Inadvertent Nuclear Yield and major release of radioactive material is based on a number of foundational safety principles, which are in turn scrutinized with respect to the probity of their underlying scientific, technical, engineering, implementation and evidential basis. This includes assurance that their implementation fully meets these principles both for normal and

abnormal (accident) environments. Of course, the ownership approach also aims to minimize the occurrence of abnormal environments and the impact of human error.

> *3. Strive for systems that are simple, easily understood, disconnected, and decoupled*

Response in warhead design: This is essentially a re-statement of point 1 above. We adopt this strategy for nuclear weapon design safety focusing on clarity, simplicity and strong independent arguments in the defence in depth approach. Safety is based on a limited number of very strong and independent principles, arguments and technical implementations. These, together with the clarity of their safety intent and quality of implementation, enable a clear and comprehensive challenging process to be undertaken in order to fully test and analyse for overall safety completeness and compliance. The principal arguments supporting the safety case should be simple and clear and should not arise as a result of a complex set of contributing arguments.

The strategy is based on the application of *a limited number* of clear strong and independent safety arguments and implemented safety systems based upon ensuring maximum resilience against the undermining of safety that can occur through added complexity, as the number of safety systems increases. Increasing the number inevitably leads to increasing complexity and increasing concerns about maintaining true independence between the implemented safety systems. Added complexity generally increases the difficulty of ensuring confidence of overall safety because of the multiplicity of potential sneak paths or fault sequences that can give rise to failure. In addition, a warhead design is confronted by constraints on available volume and mass and the need for the complementary requirement that these safety systems can be reliably removed when fully 'authorised' to do so, and such authorisation is based on a principle of uniqueness which effectively decouples it from the pre-authorisation safety argument.

6 Some Additional Factors

6.1 A Lesson from History — Predicting the Future based on the Past

In 1907, five years before he went down with his ship, the Captain of RMS Titanic, E. J. Smith, was quoted in a New York Times interview:

> *But in all my experience, I have never been in any accident… of any sort worth speaking about. I have seen but one vessel in distress in all my years at sea. I never saw a wreck and have never been wrecked nor was I ever in any predicament that threatened to end in disaster of any sort.*

The clear lesson here is that past history (or perceived past history), whilst it may be a guide to the future, is only a guide and so gives *no* assurance about the future. Confidence (false) was also based on the safety built into the Titanic, where all potential threats were deemed to have been considered and catered for. The ship was judged 'unsinkable' but in fact suffered a catastrophe on its maiden voyage. This resulted from a combination of failure to prevent the collision and an overestimate of the ship's capability to withstand such a collision. *This was an example of a 'Known — but not to the degree necessary'*. Similar arguments can be applied to the Columbia and Challenger tragedies where confidence was again based on 'past successes and inadequate technical assessment of known threats'.

Warnings from the past come in the form of failures and 'near misses' and, if acted upon, can improve the situation. However, near misses are not always recognized for what they are and as such there is no guarantee they are fully understood with regards to the future. AWE however scrutinizes issues of this nature to ensure that the fundamental reasons for their cause are fully understood and removed. We *may* take some comfort from the past-history of UK's nuclear weapons ownership where there have been no major safety events, but one cannot take this as assurance in respect to the future. Certainly, the number of successful weapon lifetimes in our history comes nowhere near to proving compliance with the exacting safety standards that we strive to comply with. We certainly take heed of any issues that have arisen and have acted upon them with high priority.

6.2 Human Factors and Independence

Humans are not exempt from failure. Assumptions of true independence in human actions, which is often used as a major redundancy element in preventing a failure, has a long history of being undermined; note the US Minot incident[5]. This type of failure has been noted in dual 'independent' checking processes where enhanced confidence is based on an assumption that true independence is assured in the process. Humans are seldom truly independent in this context. For this reason, the safety of nuclear warheads is based principally on scientific, technical and engineering principled foundations. Nevertheless. 'independent' checking and dual control, together with human error assessment, do have an important place in nuclear warhead safety in preventing and quantifying inadvertent actions. Of course, in the limit even safeguards based on technical principles are not absolutely free from human influence. For example, engineering-based safeguards need human actions for design, manufacture, inspection and maintenance for ensuring and maintaining design intent compliance.

6.3 Conservatism, Worst Case and Probability Function

The approach to nuclear warhead safety design and assessment is conservatively based. This conservatism is applied to combat any possible remnant uncertainty in occurrence probability and consequence and usually takes the form of a somewhat 'worst-case' rather than best-estimate approach to risk assessment. This less onerous (in application) approach is well understood and provides a sensible hedge against uncertainty, but will not always be usable. In some cases, 'worst-case' assumptions can be *too* conservative, and lead to difficulty in meeting exacting risk safety standards. Somewhat less conservative but still supportable statistical approaches, Figure 2, are then deemed more appropriate. This tends towards a more onerous approach (in application) and can carry with it a smaller conservative margin in protecting against uncertainty.

As noted previously, distributions often have greater evidential support near their peaks and far less in the tails, which typically characterise the low probability associated with potentially catastrophic events — Black Swans. The safety assessment of nuclear warheads sometimes puts us in these regions where we need to apply such tails of distributions with suitable caution and with a sound evidential base founded on substantial supporting technical and historical arguments. Such distributions are only accepted if detailed scrutiny provides the evidence noted above and that in turn this evidence stands

[5] An unauthorized transfer of nuclear warheads between Minot Air Force Base, North Dakota, and Barksdale Air Force Base, Louisiana, 29th — 30th August 2007.

up to independent challenge. This typically applies to assessment of explosive response to environmental challenge.

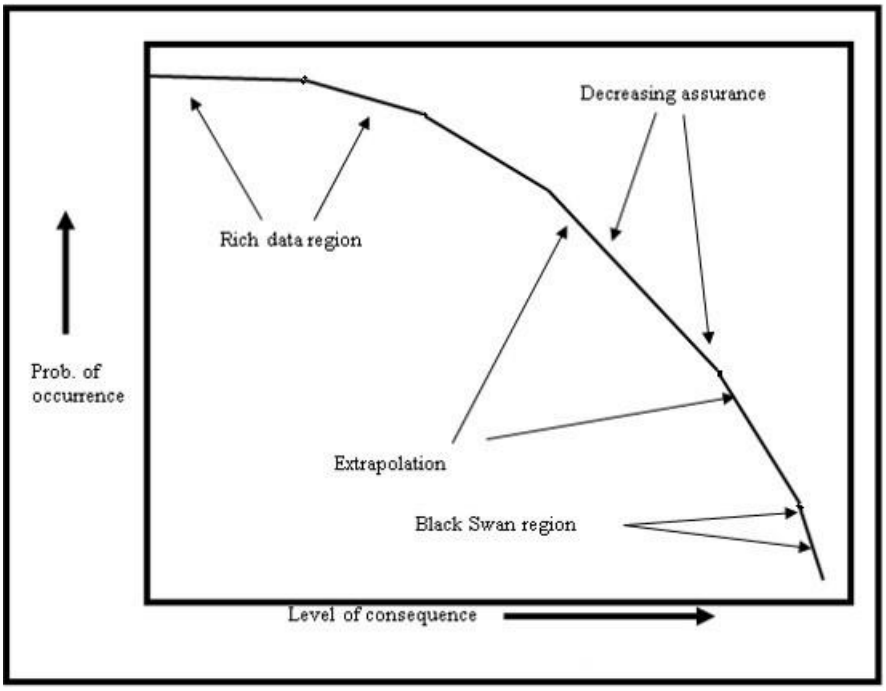

Figure 2 ~ Distribution Approach to Safety Assessment

7 The Importance of, and Achieving, Independence

The occurrence of mishaps and even Black Swans are often the result of an initiator setting off a related sequence which is not truncated. Protection against this occurrence will rely heavily on:

- A full understanding of the inventory of possible hazardous sequences leading on from any given initiator and culminating in a mishap.
- Ensuring a suitable level of independence between such events in the sequence so that the sequence is truncated before reaching the mishap stage.

Some of these aspects were covered in Jones (2017) in terms of the identification of the strength in depth of barriers or Lines of Defence (LOD) in relation to their number and strength in preventing such a mishap. In particular, the importance of independence between such LOD. The choice of LOD is based on the widest range of categories possible in order to meet the independence requirement. For example:

1. Fundamental physics or chemical properties based on the laws of nature, which directly support the LOD resilience
2. Uniqueness of response. In the sense that defeat of the LOD can only occur as a result of unique circumstances.
3. That microelectronics/firmware/software elements are clearly decoupled from the primary safety arguments.
4. Best practices engineered implementation of a concept to the extent that failure is unlikely even under challenging circumstances, e.g. accident conditions.

5. Procedure based. This relies on humans doing the correct thing at the correct time, and not doing the incorrect things at the wrong time. Examples are fully trained staff and application of the two-man checking principle.

These can be broadly categorised as having the resilience properties of: **Isolation, Incompatibility, Inoperability, Independence**. Of course, although correct procedure can have a significant influence in safety, it takes the lowest place in the hierarchy of attributes in the strength in depth argument.

The ability to deterministically make an LOD resilience case and to support the principle of LOD independence will follow this hierarchy. A structured sequence protection argument based strongly on correct LOD based procedures alone will not be acceptable. Although human LOD are least favourable in the overall choice, all LOD have some element of human involvement because of human relationship to judgement, inspection, testing, maintenance, etc.

In addition to the choice of LOD category, the overall strength in depth arguments for LOD in any particular fault sequence is enhanced by applying the following hierarchy of application for independence in as far as technical implementation allows:

1. Fundamentally different concepts.
2. Fundamentally different applications of the same concept.
3. Different engineering implementations of the same concept.
4. Application of different and differently-sourced materials in the same concepts or engineering implementation.

All of these LOD attributes form the major basis of preventing sequences propagating to the mishap conditions.

8 The Influence of Environmental Factors on Black Swan Assurance

An initiator can take the form of a design flaw (failure) or an ‘insult’. The latter represents an abnormal environment (accident) and this environment could in turn be applied simultaneously to all LOD in the strength in depth strategy. As such it is important to avoid common mode failure in the application of the LOD. General historical evidence shows that many major mishaps have indeed arisen from common cause environmental based failures of this manner. The potential for abnormal environments presents the greatest challenge in ensuring that all of the sequence paths are identified for any warhead design. In turn one must show that there is sufficient resilience to ensure that such sequences, through design and testing, are truncated under these circumstances and that there is no common mode failure. Prevention of abnormal environment occurrence is of course also a major objective in the overall safety theme and every effort possible is made to prevent/limit the potential for such insults/accidents and in turn the prevention of such environments propagating to the warhead. Nevertheless, such events cannot be totally discounted and as such the safety characteristics of a warhead design must show appropriate resilience against such threats.

9 Conclusions

Hunting for the Black Swan in the context of safety can represent a somewhat uncertain activity in the sense that one may be looking for something that does not exist. One cannot prove a negative. However, there is a compelling history of disasters which evidence the existence of such Black Swans which were not identified in advance. The level of scrutiny applied to the search for Black Swans must be related to the potential consequence of failure. Of course, for nuclear warheads we are well aware of the worst-case catastrophe represented by inadvertent nuclear yield and, as a result, have established a sound culture aimed at minimising the potential for such an occurrence. In addition, we need to provide a compelling level of evidence/assurance that the probability of such a (man-made) occurrence will be less than natural catastrophes of a similar magnitude. We do this based on an independent strength in depth approach coupled with state-of-the-art application of science, engineering and technology with all of the associated requirements set by established best practice. Such designs are then subjected to an independent conservative numerical risk assessment based on a foundation of supporting evidence to demonstrate compliance with demanding national risk standards. Both fundamental design aspects and the risk assessment are in turn subject to organisational independent assurance challenge. The latter covers the fidelity and completeness of the: logic, evidence, implementation and a search for any aspects which might have been missed or not covered in sufficient depth. Both contributions are then considered at the organisational executive decision-making level where a further level of knowledge and experience is applied. These activities go a long way to eradicating the presence of Black Swans but of course this does not guarantee that one is not present. The final layer in the defence comes from the strong organisational support to a continuing probing[6] culture which applies a 'what if' and 'expecting the unexpected' mind-set. Consequently, the campaign for assuring safety never ends. This latter contribution is based on a strong and continuing organisational commitment to engaging the best brains, which continue to probe into the various science, technology and design areas with the goal of achieving even greater depths of understanding in the cause of potential Black Swan identification and early eradication before such can manifest itself. This is exercised both through direct in-house activity and through what can be gleaned from the external world. One cannot claim that this last layer leads to perfection and at times may appear to be somewhat 'undirected' in its approach. However, it is still the best way we know of with regard to tackling the demanding challenge of Unknown Unknowns. The overall strategy is hence based on:

1. A sound approach for developing a safe product and process — ensurance.
2. A sound independent scrutinising/challenging element — independent assurance.
3. A knowledgeable and experienced executive decision-making element — final accountability.
4. An organisational commitment to applying the best brains to a continuing probing approach into all of the technology and procedural areas — the last possible hiding place of the Black Swan — safety is never finished.

Disclaimer

The contents of this document represent the views of the author and not necessarily those of AWE plc.

[6] i.e. questioning and challenging

Correspondence Address

Malcolm Jones, Room 110, AWE, Aldermaston, Berkshire RG7 4PR UK

Malcolm.Jones@awe.co.uk

Acknowledgments

The author acknowledges the helpful, interesting and informative discussions held with members of AWE and the International System Safety Society when preparing this paper.

References

Jones M. (2016). *Organisational Problems — Potential Causes — Unintentional Consequences.* In: 34th International System Safety Conference, Orlando, Florida, USA, 8 -12 August 2016. System Safety Society, Unionville.

Jones M. (2017). *Strength in Depth and Quantitative Risk Assessment in the Context of Low Frequency High Consequence Systems*. In: 35th International System Safety Conference, Albuquerque, New Mexico, 21-25 August 2017. System Safety Society, Unionville.

Nelson R. (1964). *Fate Is the Hunter*. Arcola Pictures Corporation, Los Angeles.

Perrow C. (1972). *Complex Organizations: A Critical Essay*. McGraw-Hill, New York.

Perrow C. (2007). *The next catastrophe: Reducing our vulnerabilities to natural, industrial and terrorist disasters*. Princeton University Press, New Jersey.

Smith E. J. (1907). *Captain, RMS Titanic Quotes*. Quotes.net. Retrieved 5th May 2023 from https://www.quotes.net/quote/35858.

Taleb N. N. (2008). *The Black Swan: The Impact of the Highly Improbable*. Penguin Books, London.

This collation page left blank intentionally.

ISSN 2754-1118 (Online)
ISSN 2753-6599 (Print)

About the Safety-Critical Systems eJournal

Purpose and Scope

This is the Journal of the Safety-Critical Systems Club CIC (SCSC), ISSN 2754-1118 (Online), ISSN 2753-6599 (Print). Its mission is to publish high-quality, peer-reviewed articles on the subject of systems safety.

When we talk of systems, we mean not only the platforms, but also the people and their procedures that make up the whole. Systems Safety addresses those systems, their components, and the services they are used to provide. This is not a narrow view of system safety, our scope is wide and also includes safety-related topics such as resilience, security, public health and environmental impact.

Background

When the Safety-Critical Systems Club (SCSC) was set up over thirty years ago, its objectives were to raise awareness of safety issues and to facilitate safety technology transfer. To achieve these objectives, the club organised events, such as Seminars and an annual Symposium, and published a newsletter, Safety Systems, three times a year.

The Newsletter has, in addition to news, opinion, correspondence, book reviews, and the like, also carried articles discussing current and emerging practices and standards. The length of such articles is limited to about two and a half thousand words, which does not allow an in-depth treatment. It was therefore decided to add a third string to our bow and supplement the events and newsletter with a journal containing longer papers. The journal is now published here, as the Safety-Critical Systems eJournal, and comprises at least two issues a year.

Content Sources

Sources include the outputs of SCSC working groups; solicited technical articles and topic reviews; submitted articles on new analysis techniques, discussion of standards, and industrial practice; and guidelines and lessons learned. If you wish to contribute, please see, "Information for Authors".

Types of paper include, but are not limited to:

Technical Articles: Written by practitioners and describing practical safety assurance techniques and their industrial applications.

Integration Studies: Written by practitioners reporting upon successful (or otherwise) synergies achieved in practice with other assurance domains, e.g. systems engineering, reliability/availability/maintainability engineering, resilience, human factors, security, and environment.

Position Papers: Written by, or on behalf of, Regulators, Standardisation Organisations, or other official bodies, setting out their position on a topic, e.g. the interpretation of a particular standard or regulation.

Review Articles: Papers highlighting recent developments and trends in some aspect of safety-critical systems or of their use in a particular industrial sector.

Historical Articles: Papers describing the development of safety assurance in an industrial sector; how we got to where we are today.

ISSN 2754-1118 (Online)
ISSN 2753-6599 (Print)

Perspectives: The authors' personal opinions on a subject, e.g. whether to use statistical methods in particular scenarios.

Reports: The lessons learned from incidents or the outcomes of trials with a description of scenarios, or methods, and a discussion of the results obtained.

Working Group Outputs: Written by Safety-Critical Systems Club Working Groups to include discussions, underpinning theory, or guidelines.

Copyright and Disclaimer

The author(s) of each paper shall retain copyright in their work but give the Safety-Critical Systems Club permission to publish in both on-line and printed formats. While the authors and the publishers have used reasonable endeavours to ensure that the information and guidance given in this work is correct, all parties must rely on their own skill and judgement when making use of this work and obtain professional or specialist advice before taking, or refraining from, any action on the basis of the content of this work.

Neither the authors nor the publishers make any representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability or availability with respect to such information and guidance for any purpose, and they will not be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever (including as a result of negligence) arising out of, or in connection with, the use of this work. The views and opinions expressed in this publication are those of the authors and do not necessarily reflect those of their employers, the Safety-Critical Systems Club, or other organisations.

Letters to the Editor

The editorial to the first issue of this journal said, "*You may find some of this material controversial, or you may think that it does not go far enough. Subsequent issues of this journal will have provision for readers' letters to the Editor responding to individual papers.*" Such a letter should be no more than 1000 words in length (not counting title, attribution, or references). That would take up no more than two pages of the journal. Note that a letter should ideally address a single concern with few, if any, external references.