Cyber Evaluation and Management Toolkit (CEMT): A threat-focused, model-based approach to assessing the cyberworthiness of cyber-physical systems to facilitate informed decision-making

Author:
Fowler, Stuart

Publication Date:
2025

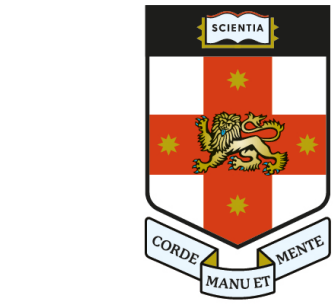
DOI:
https://doi.org/10.26190/unsworks/31202

Cyber Evaluation and Management Toolkit (CEMT): A threat-focused, model-based approach to assessing the cyberworthiness of cyber-physical systems to facilitate informed decision-making

Stuart Fowler

A thesis in fulfilment of the requirements for the degree of

Doctor of Cybersecurity



UNSW
SYDNEY

School of Computer and Systems

UNSW Canberra

May 2025

DECLARATIONS

ORIGINALITY STATEMENT

☑ I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

COPYRIGHT STATEMENT

☑ I hereby grant the University of New South Wales or its agents a non-exclusive licence to archive and to make available (including to members of the public) my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known. I acknowledge that I retain all intellectual property rights which subsist in my thesis or dissertation, such as copyright and patent rights, subject to applicable law. I also retain the right to use all or part of my thesis or dissertation in future works (such as articles or books).

For any substantial portions of copyright material used in this thesis, written permission for use has been obtained, or the copyright material is removed from the final public version of the thesis.

AUTHENTICITY STATEMENT

☑ I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis.

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in the candidate's thesis in lieu of a Chapter provided:

- The candidate contributed **greater than 50%** of the content in the publication and are the "primary author", i.e. they were responsible primarily for the planning, execution and preparation of the work for publication.
- The candidate has obtained approval to include the publication in their thesis in lieu of a Chapter from their Supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis.

☑ The candidate has declared that **their thesis has publications - either published or submitted for publication - incorporated into it in lieu of a Chapter/s. Details of these publications are provided below.**.

Publication Details #1

Full Title:	Cyber Evaluation and Management Toolkit (CEMT): Face Validity of Model-based Cybersecurity Decision Making
Authors:	Stuart Fowler, Keith Joiner and Siqi Ma
Journal or Book Name:	MPDI Systems
Volume/Page Numbers:	
Date Accepted/Published:	
Status:	published
The Candidate's Contribution to the Work:	Conceptualization, methodology, software, investigation, writing—original draft preparation, visualization
Location of the work in the thesis and/or how the work is incorporated in the thesis:	This work in incorporated as Chapter 5 - Face Validity Trials. The work is used to describe the initial trials of the CEMT and the results of those trials.

Publication Details #2

Full Title:	Assessing Cyberworthiness of Complex System Capabilities using the Cyber Evaluation and Management Toolkit (CEMT)
Authors:	Stuart Fowler, Keith Joiner and Siqi Ma
Journal or Book Name:	Computers and Security
Volume/Page Numbers:	
Date Accepted/Published:	
Status:	submitted
The Candidate's Contribution to the Work:	Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Conceptualisation.
Location of the work in the thesis and/or how the work is incorporated in the thesis:	This work is incorporated as Chapter 6 - Implementation Trials. The work is used to describe the practical implementation trials of the CEMT and the results of those trials.

Candidate's Declaration

I confirm that where I have used a publication in lieu of a chapter, the listed publication(s) above meet(s) the requirements to be included in the thesis. I also declare that I have complied with the Thesis Examination Procedure.

ABSTRACT

Ensuring the cyberworthiness of complex cyber-physical systems is a critical component of designing modern military capabilities. The requirement to ensure the mission systems are resilient to cyber attacks and able to operate safely and effectively within a contested cyber environment is widely articulated, but the path to achieving that requirement efficiently and effectively is less clear. Our initial hypothesis, based on a survey of the available literature, identified that a model-based approach that focused on detailed modelling of adversarial threats could produce assessments that were both sufficiently detailed and intuitive to decision makers, but existing model-based approaches were unable to satisfactorily meet these objectives.

A new model-based approach, the Cyber Evaluation and Management Toolkit (CEMT), was developed to address the gaps in the literature, and early prototypes of this approach were evaluated via face validity trials using generic case studies that surveyed expert respondents within the Australian Defence enterprise. Promising results to these face validity trials paved the way for field research and an implementation trial that applied the CEMT to the assessment and evaluation of the cyberworthiness of a complex military capability within the Royal Australian Navy (RAN). Stakeholders involved in the implementation trial responded to surveys to gather data on the efficiency, effectiveness and user satisfaction of our new approach as a tool for assessing the cyberworthiness of a complex cyber-physical system.

The results of the implementation trial generally supported the claim that the CEMT was both an efficient and effective tool for assessing cyberworthiness, and the overall user satisfaction of the tool was high – especially in comparison to extant approaches. The results also identified that the CEMT facilitated informed decision making with respect to cyber risk.

These results support our initial hypothesis that threat-informed, model-based approaches can be used to systematically assess the cyberworthiness of complex cyber-physical systems and present that assessment in a manner that is intuitive and understandable to decision makers. The CEMT provides a strong example of such an approach and further development of the tool could lead to an operationalised capability for assessing the cyberworthiness of complex cyber-physical systems.

CONTENTS

TABLES AND FIGURES

ACKNOWLEDGEMENTS

Completing this doctoral journey has been a profound experience. Over the course of this journey, I have become acutely aware that it was not one I could have taken alone, and I am deeply grateful for all the people who have helped me progress through this research program over the past six years.

Firstly, I would like to express my heartfelt thanks to my supervisors, especially Dr Keith Joiner. Your guidance, expertise, and unwavering support have been invaluable in shaping my research and helping me reach this milestone. Your patience and encouragement have helped me to push boundaries and strive for excellence.

Secondly, my parents have provided me with amazing opportunities throughout my life and taught me the value of ongoing education and I cannot thank them enough. Everything I have ever done and everything I have ever achieved is only possible because of the foundation they built for me during my childhood, and I remain ever grateful for the efforts and sacrifices that they made to ensure I had the best start in life. Dad, thanks for the continual support and for the role that you model. Mum, you will always be missed.

I would also like to thank all the other friends and family who became overnight editors, reading through my work to pick up all the typographical errors and grammatical oversights that slipped into those early drafts. If the involuntary

reaction my body makes when I think of carefully reading through these chapters again is anything to go by, you guys are the real heroes.

To Caitlin, whose love and understanding has been a guiding light through this venture – thank you for your endless support and your words of encouragement when the challenges seemed insurmountable, for affording me the opportunity to focus on this project and for giving me the strength to see it through to the end. This truly would not be possible without you cheering me on, and while I think even a lifetime of gratitude won't be enough to repay you for everything you do for me, my gratitude is something you will undoubtedly have as we continue to share a lifetime of happiness together.

Finally, I would like to dedicate this whole endeavour to my children, Emily, Ben and Lucy. I hope that this inspires you to always work toward something bigger and demonstrates the value of lifelong learning and continual education.

PUBLICATIONS AND PRESENTATIONS

A list of any relevant publications or presentations arising from the writing of the thesis should also be included.

JOURNAL ARTICLES

1. Fowler, S.; Joiner, K.; Ma, S. *Cyber Evaluation and Management Toolkit (CEMT): Face Validity of Model-Based Cybersecurity Decision Making*. *Systems* 2024*,* 12*,* 238. https://doi.org/10.3390/systems12070238
2. Fowler, S.; Joiner, K.; Ma, S. Assessing Cyberworthiness of Complex System Capabilities using the Cyber Evaluation and Management Toolkit (CEMT), *preprint* [submitted]

CONFERENCE PAPERS

1. Fowler, S.; Sitnikova, E. *Toward a framework for assessing the cyber-worthiness of complex mission critical systems*. In Proceedings of the 2019 Military Communications and Information Systems Conference (MilCIS), Canberra, Australia, 12-14 November 2019.

CONFERENCE TUTORIALS

1. Fowler, S.; Joiner, K.; Sitnikova, E.; Ma, S. *Cyberworthiness Evaluation and Management Toolkit (CEMT): A model-based approach to*

cyberworthiness assessments. Systems Engineering Test & Evaluation Conference, Canberra, Australia, 12 September 2022

INDUSTRY PRESENTATIONS

1. Fowler, S. *Cyberworthiness Assessments using the Cyber Evaluation and Management Toolkit*. Australian Energy Intel Group (EIG). Sydney, Australia, 31 August 2022. *Invited by Energy Queensland and Ausgrid.*
2. Fowler, S. *Cyber Authorisation Modernisation using the Cyber Evaluation and Management Toolkit (CEMT)*. Australian Department of Defence Cyberworthiness Community of Interest. Canberra, Australia, 7 March 2024. *Invited by Australian Department of Defence Cyberworthiness Coordination Office.*
3. Fowler, S. *Assessing the cyberworthiness of complex system capabilities using the Cyber Evaluation and Management Toolkit (CEMT)*. Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Cyber Pavillion. Orlando, Florida, USA, 6 December 2024. *Invited by Command Post Technologies.*
4. Fowler, S. *Assessing the cyberworthiness of complex system capabilities using the Cyber Evaluation and Management Toolkit (CEMT)*. US DoD Joint Cyber Community of Practice. [Virtual], 22 January 2025. *Invited by US Office of the Undersecretary of Defence (OUSD) Research and Engineering (R&E).*

CHAPTER 1 INTRODUCTION

1.1 BACKGROUND

This research revolves around the emerging efforts to appropriately secure cyber-physical systems – those systems that combine general-purpose computing devices, physical processes and networked communications – against modern cybersecurity threats. To understand the background to cybersecurity in cyber-physical systems, we must first understand the history of technology convergence that led to the proliferation of cyber-physical systems, as well as an understanding of how cybersecurity threats and the associated cybersecurity industry has evolved over the same period.

1.1.1 Technology Convergence

Technology convergence refers to the coalescence of distinct technologies into integrated systems that perform a holistic function. In the context of cyber-physical systems, technology convergence is used to describe the ongoing process of Information Technology (IT) systems merging with Operational Technology (OT) to form a combination of embedded systems, physical processes, computing devices and networked communications. To understand this background for cyber-physical systems, we will briefly explore the history of IT and OT systems.

The birth of modern Information Technology (IT) networks largely begins with the development of packet switching technology and its first applications within ARPANET in the late 1960s and early 1970s [1], which provided the foundation for distributed wide-area networks. Around the same time, the invention of the integrated electronic circuit facilitated the transition from mainframes to minicomputers [2], which was shortly followed by the creation of the microprocessor that drove the phenomenon of personal computing [3]. These advances in computer hardware, and the exponential costs savings that they brought with them, led to an explosion in the number of computing systems around the world. In parallel, the first versions of the now ubiquitous Transmission Control Protocol / Internet Protocol (TCP/IP) suite were being developed to expand ARPANET, with the familiar version 4 of the protocol suite eventually published in 1980 [4].

These protocols supported the interconnection of the rapidly expanding number of computers, giving birth to modern internetworking. Over time, we added the world wide web, email, instant messaging, voice calls, video calls, online shopping services and eventually social networking services, until this global network of computing devices resembled the critical backbone of society that we know as the Internet. The introduction of mobile data networks and the advancement of battery technologies expanded wide-area networks to include mobile computing devices, and the age of ubiquitous computing was upon us. In modern society, IT systems are critical components of almost every facet of the economy and have become deeply intertwined with our daily lives – constant interconnectedness via IT systems is considered the norm, and 'switching off' is seen as a temporary deviation from the default state of omnipresent internetworking.

Operational Technology (OT) includes a broad range of embedded control systems and devices that detect or change physical processes [5]. While microprocessors drove the advancement of general-purpose computing in the realm of Information Technology, the closely related microcontroller and the specialised embedded systems built using those microcontrollers were the fundamental component of early Operational Technology. As they became cheaper, these embedded digital control systems replaced the analogue mechanical systems that controlled early machinery and other automated physical systems. These embedded digital control systems monitored a physical process using sensors and controlled that process through actuators [6].

Over time, these individual control systems that controlled individual physical processes, were interconnected via communication networks using different networking protocols to form what we would now consider an Operational Technology (OT) network. Examples of modern OT systems include:

- Industrial Control Systems (ICS), including
    - Supervisory Control and Data Acquisition (SCADA) systems
    - Distributed Control Systems (DCS)
    - Programmable Logic Controllers (PLCs)
- Building Automation Systems (BAS)
- Physical Access Control Systems (PACSs)

In a traditional OT system, the communication networks that interconnected these components were not packet-switched networks, instead utilising dedicated circuits or communication busses to achieve networked communication. The OT systems themselves tend to be more specialised and more proprietary than IT systems which tend to be more general-purpose and built on open standards.

However, more recently there has been a continuing convergence between IT and OT systems. Increasingly, the TCP/IP protocol suites that dominate internetworking of IT systems has been leveraged for the networked communications of OT systems and this has facilitated the progressive integration of these OT networks with the wider Internet. Through the convergence of IT and OT systems, the concept of a cyber-physical system was

born – a system which incorporates general-purpose computing devices, physical processes and networked communications. By merging these traditional information technologies into traditional operational technology systems, the cyber-physical system has been exposed to both significant new functionality and significant new risks.

1.1.2 Cybersecurity Evolution

While IT and OT systems were themselves developing, so were the cybersecurity threats against them, leading to an escalating battle between those trying to attack the systems and those defending them. This battle was set against a backdrop of the increasing interconnectivity discussed in the previous section.

The first significant milestone in the evolution of cybersecurity occurred in 1971 with the *Creeper* virus, which while harmless was able to replicate across ARPANET, showcasing the vulnerability of early networked systems [7]. In the 1980s, cybersecurity threats began targeting personal computers and their networks, with the *Brain* virus being a notable example from 1986. This virus infected floppy disks and demonstrated that malware could spread through physical media. The *Morris* worm, created in 1988 was one of the first worms capable of spreading through vulnerabilities in the services that facilitate network connections, and it ended up compromising an estimated 10% of the fledgling Internet at the time [8]. Given this significant impact, the Morris worm illustrated the potential consequences of cybersecurity threats and forced a greater emphasis on the need for cybersecurity defences.

The 2000s saw the proliferation of cybersecurity threats delivered via email systems. The *ILOVEYOU* worm in 2000 spread by enticing recipients to open malicious attachments disguised as love letters and the *Mydoom* worm in 2004 also leveraged malicious attachments to infect over 500,000 machines within a week [9]. These viruses demonstrated that even when network services were secured against attackers, those attackers could use social engineering to undermine those defences.

In the early days of OT systems, they were generally isolated from the Internet and, because of this, were considered immune to cyber threats. However, a pivotal moment in OT cybersecurity occurred in 2010 with the deployment of *Stuxnet*, a highly sophisticated worm that targeted Iran's nuclear enrichment facilities. *Stuxnet* highlighted that cyber threats could cause physical damage by sabotaging industrial control systems, making the dawn of a new era in OT cybersecurity. Stuxnet also leveraged a well-known vulnerability used by the *Conficker* worm, which was identified and patched in 2008, confirming that these OT systems were significantly behind IT systems from a cybersecurity maturity perspective [10].

The late 2010s included a surge in ransomware attacks and nation-state cyber operations. The *WannaCry* ransomware attack in 2017 affected hundreds of thousands of systems globally, including critical infrastructure like hospitals [11]. In the same year, the *NotPetya* virus crippled shipping infrastructure and paralysed several major organisations in one of the most devastating cyber

attacks in history. The *NotPetya* virus relied on an exploit known as *EternalBlue*, which was created by the US National Security Agency but leaked in a security breach earlier in 2017 [12]. These attacks highlighted the increasing cyber risk to critical infrastructure as the convergence of IT and OT systems lead to an expanded attack surface against these once-isolated OT systems.

The current era of cybersecurity threats is marked by Advanced Persistent Threats (APTs) and state-sponsored actors. The convergence of IT and OT systems, and the subsequent emergence of cyber-physical systems, has blurred the line between digital and physical security, requiring specialised cybersecurity solutions that address both traditional and emergent risks. The ability to adequately protect cyber-physical systems against cyber threats remains a pressing concern as attackers continue to innovate.

1.1.3 A New Normal

In this threat environment, cybersecurity threats need to become a key consideration of the design, deployment and operation of cyber-physical systems. There is a clear need to ensure that cyber-physical systems are resistant and resilient to these cybersecurity threats, and it is tempting to look toward the well-established IT cybersecurity industry that has evolved alongside this threat for solutions to this emerging challenge. However, the existence of operational technologies within cyber-physical systems demands holistic security approaches that do not just focus on the IT components embedded within the cyber-physical system.

It is not clear that the hard learnt lessons that cybersecurity practitioners have gathered from decades securing IT systems will be transferable directly to cyber-physical systems. Similarly, it is not clear that the equivalent security standards that have been developed for industrial control systems and other operational technologies [5] will satisfy the cybersecurity needs of cyber-physical systems either. This is because in a cyber-physical system which converges both IT and OT, a cyber risk within the IT components can cause a consequence within the OT components and vice versa.

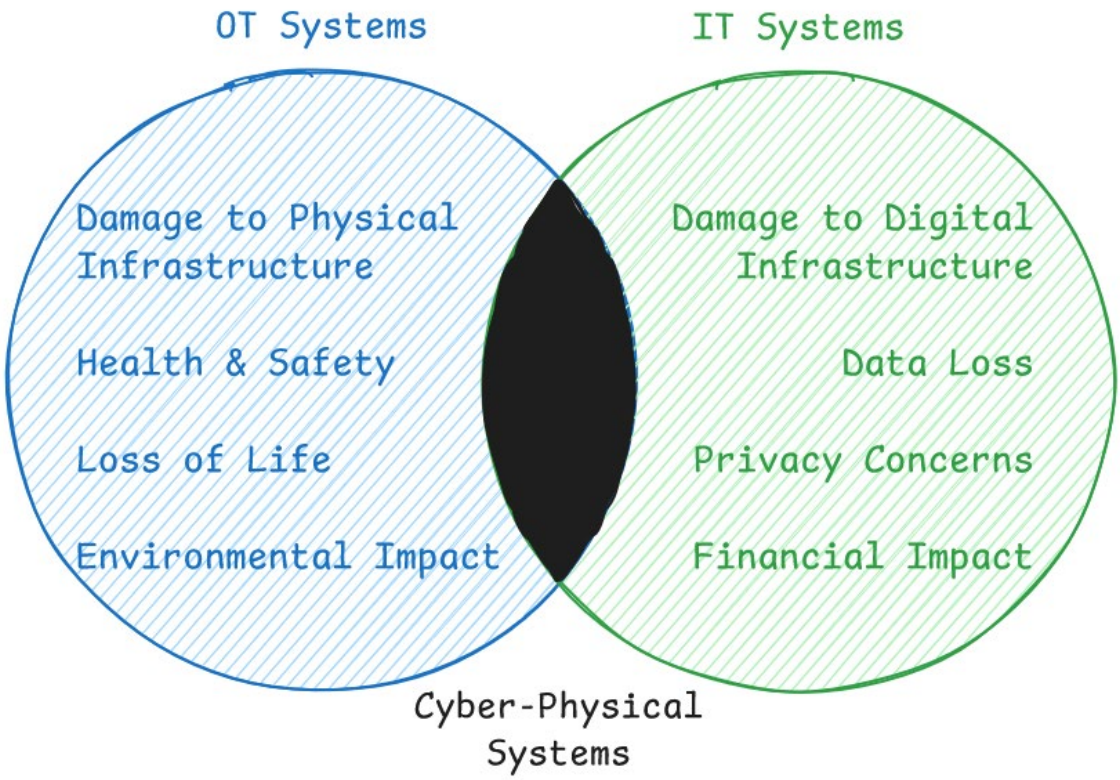


Figure 1-1: *Consequences of OT versus IT cyber-attacks. Adapted from [10].*

This difference in consequences between a compromised IT system vs a compromised OT system is depicted in Figure 1-1. The primary difference is that

the consequence of an IT system compromise is rarely a direct human safety impact, whereas this direct impact on human safety is a common consequence of an OT system failure or compromise.

As a result of this, safety analyses of cyber-physical systems, either through workplace health and safety regulations or system safety standards, are looking at cyber-attacks against the cyber-physical system as a cause to preventable safety mishaps. This brings with it a demand for increased focus on availability and integrity of the components in the cyber-physical system and higher levels of engineering assurance behind any risk-based cybersecurity decisions.

Equally, the consequences of a compromised OT system are rarely confidentiality impacts, while confidentiality loss is a key consequence in IT system compromise. Thus, in a cyber-physical system we also need to ensure that the OT components are not going to be the initial access point for an attack against confidential data in the IT components.

It is only through a holistic cybersecurity approach, tracking the threat through both the IT and OT components, that we can adequately appreciate, understand and mitigate the cyber risk. Cyber-physical systems and demands for more intelligent systems resulting in increasing interconnectivity are the new normal for the design of converged IT/OT systems, and our cybersecurity analysis techniques need to keep up. As our reliance on interconnected systems grows, so too must our commitment to protecting them.

1.2 MOTIVATION

The motivation behind this research activity comes from the perceived insufficiency of contemporary approaches to provide adequate cyber assessment of complex cyber-physical systems. With the rapid growth of an internet-connected economy driving an understandable focus of the cybersecurity industry on globally interconnected information and communications technology (ICT) systems, the best-practice contemporary approaches have been developed with this context in mind. However, not all critical systems that are the target of malicious cyber-attacks fit this mould, particularly when we consider cyber-physical systems.

1.2.1 Contemporary Approaches

The focus of this section will be on contemporary approaches mandated by government regulation, as these regulations typically bind government agencies and become pervasive within industry that deals with those government departments. Coverage will be provided of both the Australian regulatory framework and that of the United States of America.

The primary regulatory mechanism for the security of technology systems within the Australian Government is the Protective Security Principles Framework (PSPF) [13], which "*prescribes what Australian Government entities must do to protect their people, information and resources, both domestically and internationally*" [13]. Section 13 of the PSPF provides the requirements for

security technology systems, both information technology and operational technology, and requires that the controls within the Information Security Manual (ISM) [14] are applied on a risk-based approach, with the risk level accepted by the Chief Information Security Officer or equivalent accountable authority. Section 14 of the PSPF also requires that agencies implement the Essential Eight strategies to a minimum of Maturity Level Two [15].

The risk-based approach outlined in the ISM, as mandated in the PSPF, requires the following six steps to be followed:

1. *Define the system* – to determine the business criticality of the system based on the impact if the system were to be compromised.
2. *Select controls* – from the baseline controls provided in the ISM based on the desired security objectives.
3. *Implement controls* – that were selected for the system and its operating environment.
4. *Assess controls* – in the systems operating environment to determine if they have been implemented correctly.
5. *Authorise the system* – based on the acceptance of security risks associated with its operation.
6. *Monitor the system* – including the associated cyber threat, risks and controls on an ongoing basis.

The ISM contains hundreds of specific security controls designed to mitigate security threats within information technology systems and operational

technology systems. Importantly, the ISM process is not a compliance-based paradigm – there is no mandatory set of control which must be met. Rather, a control baseline of relevant controls is identified based on the system criticality and then assessed for implementation and effectiveness within the system. Security controls that are not implemented effectively are analysed to determine whether the ineffective implementation of the control increases the inherent cyber risk of operating the system. The resultant residual risk is then presented to a decision maker for acceptance – a hallmark of a risk-based paradigm.

The Essential Eight are the eight highest priority cyber mitigation strategies developed by the Australian Signals Directorate (ASD), and "has been designed to protect organisations' internet-connected information technology networks" [16]. These eight strategies are:

1. *Patch Applications*.
2. *Patch Operating Systems*.
3. *Multi-Factor Authentication*.
4. *Restrict Administrative Privilege*.
5. *Application Control*.
6. *Restrict Microsoft Office Macros.*
7. *User Application Hardening*.
8. *Regular Backups*.

The Essential Eight includes a maturity model, where each of the eight strategies can be given a maturity rating between 0 and 3, where 0 aligns with a poor or

non-existent implementation of the strategy and 3 aligns with a strong and comprehensive implementation of the strategy. While the Essential Eight is fundamentally a control baseline and could be implemented in either a compliance-based or risk-based paradigm, the mandate in the PSPF to achieve Maturity Level 2 means that, at least within the scope of Australian Government entities, the Essential Eight is implemented in a compliance-based paradigm. However, the PSPF itself does acknowledge that the Essential Eight may not be applicable outside the scope of internet-connected IT networks:

> *"While the principles behind the Essential Eight may be applied to enterprise mobility and operational technology networks, it was not designed for such purposes and alternative mitigation strategies may be more appropriate to defend against unique cyber threats to these environments." [13]*

This indicates that while compliance with the Essential Eight is mandated, it is understood that these controls won't apply to every technology system. This provides some leeway in the compliance-based paradigm that the mandate introduces, allowing for common sense to be applied where that compliance-based paradigm would lead to inefficient and ineffective cyber mitigation efforts.

Overall, this Australian Government regulatory approach to cybersecurity is primarily a control-focused, risk-based approach, which relies on a centralised set of cybersecurity controls derived from industry best-practice. Controls are selected based on the criticality of the system being assessed and a residual risk

is determined based on the level to which these controls are implemented. There is no consistent or defined methodology for how this implementation state is translated into a residual risk, and that is left up to individual assessment entities to decide. Ultimately, this residual risk must be accepted by a sufficiently accountable officer within the organisation operating the system.

The equivalent regulatory construct in the United States of America originates in the Federal Information Security Modernization Act of 2014 (FISMA) [17], which explicitly emphasises a risk-based paradigm to achieve cost-effective security. FISMA requires US federal agencies and their contractors to provide information security mechanisms commensurate with the risk and criticality of a given technology system. US federal agencies use the Risk Management Framework (RMF) [18] developed by the National Institute for Standards and Technology (NIST) to demonstrate compliance with the risk-based cybersecurity mandate of FISMA.

The NIST RMF provides a seven-step process for managing cybersecurity risks within organisations and the systems they operate:

1. *Prepare* – essential activities to prepare the organisation to manage security and privacy risks.
2. *Categorise* – categorise the system and information processed, stored and transmitted based on an impact analysis.
3. *Select* – select the set of controls to protect the system based on risk assessments.

4. *Implement* – implement the controls and document how controls are deployed.
5. *Assess* – assess to determine if controls are in place, operating as intended, and producing the desired results.
6. Authorise – senior official makes a risk-based decision to authorise the system to operate
7. *Monitor* – continuously monitor control implementation and risks to the system.

The control baseline used in the RMF process is provided by NIST 800-53: Security and Privacy Controls for Information Systems and Organizations [19], which provides a catalogue of security controls to protect organisations and the systems they operate.

The US process is very similar to the Australian process, which is unsurprising because the risk management process used by the Australian ISM draws from the US RMF [14]. That is to say that the two processes do not just appear similar, but they were purposely designed to be similar.

The generalised approach of selecting a baseline of security controls from an existing best-practice control set based on the criticality of the system being designed, implementing those controls, independently auditing the effectiveness of that implementation and then presenting a residual risk to an accountable authority for acceptance is presented as the contemporary approach for the purpose of describing the motivation for this research.

1.2.2 Suitability

There are two primary issues with the contemporary approaches outlined in Section 1.2.1 that could be improved upon:

1. Applicability to cyber-physical systems.
2. Suitability of the risk determination in cyber-physical systems.

The applicability of contemporary approaches to cyber-physical systems is degraded by the specificity of the control set. One of the paradoxes of a cybersecurity assessment on cyber-physical systems is that control implementation is often required at a very detailed level, but risk needs to be understood from a holistic, integrated perspective. If controls in a control-focused approach are defined as being implemented by components of a particular complexity – for example, by a computer operating system or by a single hardware component – then they will be applied at that level during the implementation phase and audited at that level during the assessment phase.

The typical answer to the need to achieve coverage over a complex system-of-systems is that it is decomposed into smaller, more manageable systems, which then each have the contemporary approaches applied to them on a system-by-system basis. While this is often appropriate for the control implementation steps, it is not possible to sufficiently determine the criticality of one of those systems in isolation in the earlier steps of the contemporary approaches, nor is it possible to sensibly assess the residual risk without considering the complex system-of-

systems as an integrated whole. Threats and risks that impact the emergent functionality produced by the integrated system-of-systems can be missed or misunderstood in this system-by-system approach.

This leads into the second identified issue with the contemporary approaches, which is that the specific methodology for determining risk levels is undefined in policy and left up to the assessing organisations to define. This is not necessarily a shortcoming of the contemporary approaches, because this flexibility allows for a broader applicability of the overall policy approach, but it does raise concerns around the suitability for cyber-physical systems. The prevailing methodology for determining this risk in the contemporary approaches is to have cybersecurity experts review the results of the assessment steps and use their expert intuition, based upon a knowledge of the purpose of particularly controls and an understanding of the cyber threat environment, to produce a likelihood, consequence and a risk of operating the system being evaluated. This consideration of the threat environment and threat sequence is usually implicit, occurring within the mental models that drive the expert intuition.

There are many ways in which this intuition could be misguided, but even if the intuition is accurate it is still unsuitable for cyber-physical systems because of the lack of explainability, reviewability and resultant assurance in the risk assessment and risk determination process. Expert intuition is difficult to explain to the non-experts who are being asked to make risk acceptance decisions. When making risk decisions that are based on this expert intuition, decision makers may not be

able to understand or appreciate the risk analysis and assumptions that drive the underlying expert intuition. Without the ability to critically review the analysis, it is hard to demonstrate assurance in the results of the assessment.

One of the reasons why assurance in cyber-physical systems is so important is because of the potential safety consequences if those systems were to be compromised by a cyber-attack. We can look to how safety risk assessments are conducted more generally to identify potential methods for achieving the requisite level of assurance over cyber risk assessments for cyber-physical systems. Worthiness regulations are one common approach for providing this assurance around safety in the development of cyber-physical systems. Airworthiness regulations and governance [20, 21] guide the development of aircraft, and this has spawned adjacent regulatory fields of seaworthiness [22] and landworthiness [23] being defined and enforced for their respective domains.

These worthiness systems rely on a rigorous analysis process that takes a structured approach to the safety assessment of cyber-physical systems. The rigour is expressed through explicit modelling of the causes and events preceding a potential safety incident and clear traceability between those potential events and the controls that are in place to mitigate the likelihood of those events occurring. This provides an explainable analysis process that can be peer reviewed by other experts and understood by decision makers. An equivalent process that focused on rigorous modelling of cyber threats and traceability to security controls may be able to achieve similar assurance outcomes.

Placing as much importance on rigorous threat modelling as the contemporary approaches place on identifying and assessing controls from their defined control sets, would shift the assessment approach from a control-focused one to a threat-focused one. A threat-focused approach where a control baseline is explicitly derived from the threats to a system would not only help to provide the assurance discussed above, but it would also help to address the suitability concerns for cyber-physical systems. An explicit threat model would identify which controls were relevant to the system threat environment and could provide coverage of the emergent functionality of the complex system-of-systems.

1.2.3 Importance

The final component required to explain the motivation for this research activity is an explanation of why it is important to address the concerns around the suitability of contemporary approaches for cyber risk assessment when applied to cyber-physical systems. Why does it matter if these contemporary approaches are not suited to cyber-physical systems?

Firstly, cyber-physical systems have some ability to interact with the physical world and, because of this, a loss of availability or integrity of the system can lead to direct safety outcomes. We trust these systems to perform critical functions within society, and as designers of those systems, we must earn that trust through assurance mechanisms, much like trust in the safety of complex cyber-physical systems was earned through airworthiness, seaworthiness and similar. We need an equivalent cyberworthiness regime within the design of cyber-physical

systems, ensuring that we can trust that the control systems and interlocks we rely upon to keep us safe are not vulnerable to a cyber-attack.

Secondly, the domain of cybersecurity is inherently adversarial, and the threat landscape is constantly changing and deliberately concealed. Even as cybersecurity experts, we do not have perfect information or understanding of the risks associated with operating a capability, and in that environment, it is critical that risk assessments can be presented in a manner that captures that uncertainty and highlights the level of confidence we have in each contributing factor to the overall likelihood and consequence of a risk being realised.

Simultaneously, we need to acknowledge that in an adversarial environment, even theoretically perfect security is irrelevant if it isn't timely or financially feasible. Rapid assessments are sometimes just as crucial as comprehensive assessments, but speed and accuracy are not always mutually exclusive. There have been attempts to increase the reusability of cyber risk assessments to concurrently increase both the speed and completeness of those assessments. One clear example of this is the development of structured languages to capture information developed through the contemporary control-focused approaches such as the Open Security Controls Assessment Language (OSCAL) [24]. OSCAL allows control baselines, such as NIST 800-53 and the ISM discussed in Section 1.2.1, to be expressed in a machine-readable format, and allows the assessments against those control baselines to also be expressed in that machine-readable format. Components and products that have been assessed

in one system can have their relevant data imported into an assessment for a second system that also includes that component or product.

This object-oriented approach to assessments in control-focused assessments can lead to efficiency through automation and reduction of duplicate assessment effort. We need to find a way to apply these same concepts to threat-focused assessments, to facilitate efficient assessments by standing on the shoulders of those who have already done similar work. One potential avenue for doing this is with model-based approaches, which use structured language to describe complex systems in a machine-readable format. These model-based approaches have been found to provide increased rigour, increased traceability, increased capacity for reuse, increased efficiency and improved transparency [25] when applied to the design of complex systems. The application of these model-based approaches provides a potential avenue for developing efficient and reusable threat-focused cyber risk assessments.

We can also look towards reports from critical industry to understand the importance of looking for improvements on the contemporary approaches to cybersecurity risk assessment. An audit into the Australian Department of Defence (AusDoD), which operates under the PSPF regulatory framework outlined in Section 1.2.1, by the Australian National Audit Office (ANAO) in 2024 found that less than 3% of systems within the department had been appropriately assessed and authorised from a cybersecurity perspective, and that it was taking an average of 285 days to complete each assessment [26]. While the purpose of

this audit by the ANAO was not necessarily to identify novel solutions to the identified issues, the audit does make it apparent that the contemporary approaches are not meeting the needs of the AusDoD.

The United States Department of Navy (DON), which operates under the FISMA regulatory framework outlined in Section 1.2.1, has also explicitly flagged a desire to move beyond a focus on compliance with controls toward a threat-focused approach that supports rapid, relevant assessments:

> *"The DON needs a new approach to cybersecurity that goes beyond compliance because our over-reliance on compliance has resulted in insecure systems, which jeopardize the missions these systems support. Instead of a compliance mindset, the DON will shift to Cyber Ready, where the right to operate is earned and managed every day. The DON will make this transition by adhering to an operationally relevant, threat-informed process that affordably reduces risk and produces capabilities that remain secure after they have been delivered at speed." [27]*

Clearly, there is a need and a desire to improve upon contemporary cybersecurity risk management approaches, particularly in the Defence sector. This underscores the importance of investigating efficient, threat-focused approaches and adds further weight to the overall motivation for pursuing research in this area.

1.3 DEFINITIONS

To derive a research goal from the motivation in Section 1.2, five key terms that were casually introduced in the motivation require a more rigorous definition:

1. Cyber-Physical Systems;
2. Cyberworthiness;
3. Threat-Focused;
4. Model-Based; and
5. Informed Decision Making.

In the following section, each of these key terms are defined in the context of the motivation, to better describe the central themes of a potential research goal.

1.3.1 Cyber-Physical Systems

Cyber-physical systems (CPS) are broadly defined by the American National Institute of Standards and Technology (NIST) as below:

> *"Cyber-physical systems (CPS) are smart systems that include engineered interacting networks of physical and computational components." [28]*

This broad definition from NIST can be further contextualised by the identification of some common features of CPS, specifically that they include physical processes that can be digitised with sensors, the data from those sensors can be processed and communicated to provide information related to the physical

processes and that information can then be used as the basis for decisions that can intervene in the physical processes to produce an effect [29]. While this provides some additional context, this definition remains somewhat vague and far-reaching as it appears to encompass any embedded control system.

Pathan adds that a CPS is *"a kind of cyberspace based on several interconnected physical devices, each of which is run by sophisticated embedded systems and software,"* [30] which suggests that an embedded system must reach a level of sophistication and scale to be considered a CPS. However, the level of interconnectedness between the individual subsystems within a system is also an important discriminator in defining a CPS, as articulated by Nand:

> *"Cyber-Physical Systems (CPS) is the interconnection of the virtual or cyber and the physical system. It is realized by combining three well-known technologies namely 'Embedded Systems', 'Sensors and Actuators' and 'Network and Communication System' ... In CPS the physical process and information processing are so tightly connected that it is hard to distinguish the individual contribution of each process from the output." [31]*

Ultimately, a Cyber-Physical System is conceptualised as a system-of-embedded-systems which are interconnected by and integrated with a modern computer system that orchestrates the behaviour of the embedded systems. This is a confluence of those traditionally separate systems thought of as Information Technology (IT) and Operational Technology (OT).

Of course, the reasoning and justification behind this classification of cyber-physical systems as distinct from a traditional IT or OT systems is not just because of the overlap between those two technologies. The context in which CPS are used and operated within the real world present unique challenges that demand specific focus, as explained in Rawat et al.:

> *One of the difficulties is that most of these systems are safety critical. The failure of a cell phone may not be considered catastrophic, but the failure of a car-braking system or a pacemaker can be a matter of life or death. Another difficulty is that CPS are often very large. A single modern car may have dozens of computers on it, running millions of lines of code. The sheer size of such a piece of software makes it hard to understand, maintain, and have confidence in. [32]*

Within the context of this research, a cyber-physical system is a complex system which blends IT and OT components in a symbiotic manner to produce an emergent functionality that impacts the physical world. The cybersecurity pedigree of cyber-physical systems is particularly important because of the potential impact that a loss of availability or integrity of the cyber-physical system could have on human life. The focus on cyber-physical systems within this research reflects the intention of producing a robust assurance methodology that provides sufficient rigour to underpin risk decisions that may foreseeably impact human safety.

1.3.2 Cyberworthiness

Cyberworthiness is a term that was first coined within the Australian Department of Defence [33,34], and was created to evoke the existing worthiness concepts traditionally associated with continued safety and functionality of a capability as a layer of rigorous assurance within the cyber domain. It is defined as:

> *"The desired outcome of a range of policy and assurance activities that allow the operation of Defence platforms, systems and networks in a contested cyber environment. It is a pragmatic, outcome-focused approach designed to ensure all Defence capabilities are fit-for-purpose against cyber threats." [35]*

Cyberworthiness is used in the context of this research to describe the attribute of a cyber-physical systems that is being assessed and evaluated. One measure of success within the research is therefore the level of confidence with which the cyberworthiness of a system can be determined.

Importantly, the definition of cyberworthiness included above also references the policy and assurance activities that underpin the cyberworthiness attribute of a cyber-physical system. This clearly articulates the importance of not just the end state of the system, but also the level of assurance that is achieved due to the rigorousness of the assessment process.

However, the second sentence of the definition also explicitly states that cyberworthiness is both pragmatic and outcome-focused, which means that while

the process is important, it cannot be paramount. The efficiency of the assurance activities is also critical to achieving cyberworthiness of a cyber-physical system.

Ultimately, it is this balance between rigour and efficiency, between assurance and pragmatism, that drives the use of cyberworthiness in a research goal. Any assurance activity underpinning cyberworthiness must be simultaneously rigorous and pragmatic – that is, rigorous enough to ensure cyber risk is adequately understood, but pragmatic enough that it can be feasibly applied within the real-world constraints imposed by the necessary allocation of finite resources within an adversarial domain.

1.3.3 Threat-Focused

Typically, cybersecurity assessments have three fundamental factors, and it is the interactions between these factors that are analysed and evaluated to determine the overall cybersecurity risk to a given system. These three factors are:

1. Vulnerability – a weakness within the system that can be exploited by a threat to impact the functionality of the system;
2. Threat – a malicious event or action that exploits a vulnerability to impact the functionality of the system; and
3. Control – a technical, procedural or administrative measure that is implemented within the system to address one or more threats by mitigating a vulnerability.

The relationship between the system being assessed and these three fundamental factors is shown in Figure 1-2.

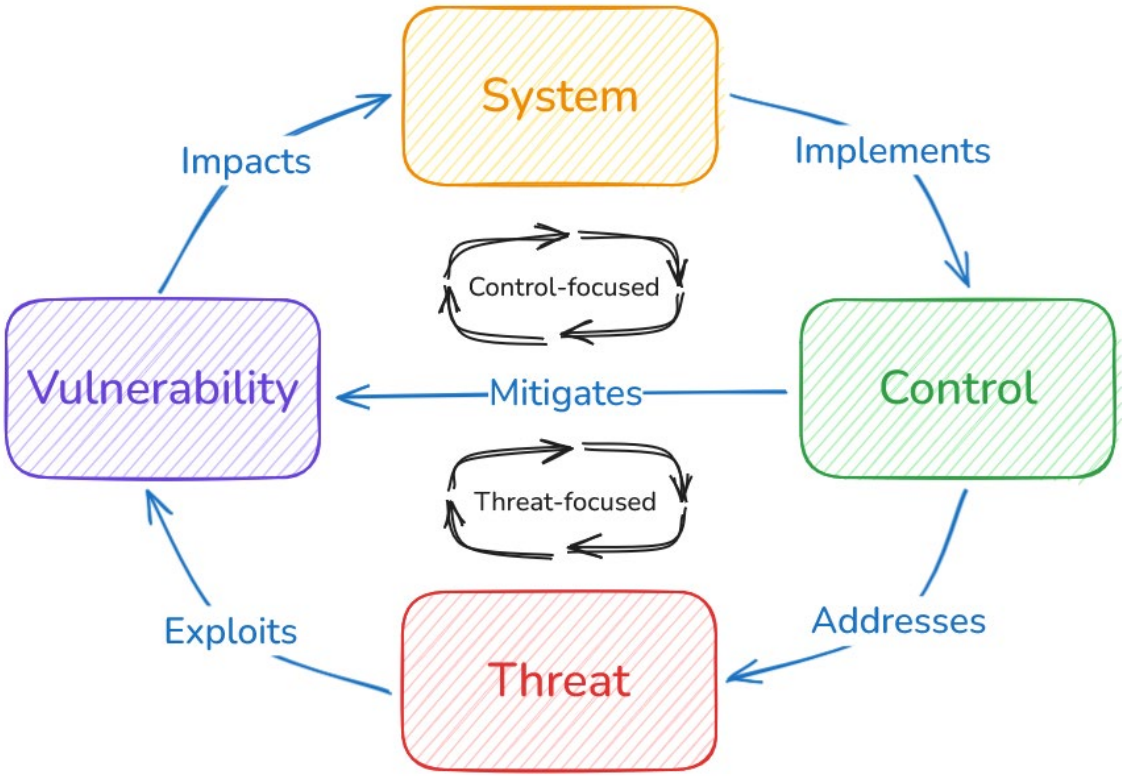


Figure 1-2: *The fundamental factors of cybersecurity risk assessment*

The likelihood of a cybersecurity risk is dependent on the probability that a *Threat* can exploit a *Vulnerability* to impact the *System* given the *Controls* that are in place. The consequence of that cybersecurity risk is dependent on the magnitude of the impact on the system. Together, these likelihood and consequence ratings are combined to determine the resultant cybersecurity risk.

A threat-focused assessment approach places primacy on the *Threat* factor. Starting from the threat, the analysis identifies any vulnerabilities within the system that could be exploited by those threats and a control baseline for the

system is derived to mitigate relevant vulnerabilities and ultimately address the threats that could exploit those vulnerabilities.

A control-focused assessment approach, however, places primacy on the *Control* factor. Rather than deriving a set of controls from the threats to a system, an existing control baseline is selected from industry best-practice and applied to a system to mitigate potential vulnerabilities in the system. The selected control baseline is typically determined by attributes of the system, such as the sensitivity/criticality of the system, or the nature of the system, such as whether it is an IT system, an OT system or cyber-physical system.

Both threat-focused and control-focused approaches can be used in either a compliance-based paradigm, where the control baselines are mandated and systems are judged on their level of compliance with the baseline, or in a risk-based paradigm, where the likelihood and consequence of an undesirable outcome is assessed, and systems are judged on the resultant residual risk.

Importantly, to generate the industry best-practice control baselines that the control-focused assessment relies upon, a threat-focused assessment must first be performed. From this perspective, the threat-focused assessment can be seen as a more ‘first principles’ approach, while the control-focused assessment can be seen as a less intensive approach that can leverage the expertise captured within industry best-practice.

Generally, this results in a control-focused assessment requiring less time investment from resources with a lower level of cyber expertise. However, control-focused assessments are premised on the assessment being performed on a system that has a similar threat profile to the generic system used by industry experts to derive the best-practice control baseline. When applied outside of that context, controls may exist in that best-practice baseline that are not relevant to the system being assessed, or the system may be subject to threats which were not considered relevant by those industry experts, leading to missing controls.

Additionally, when used within a risk-based paradigm, the threat to a system is a key component in determining the likelihood of a cybersecurity risk. Without the threats being specifically enumerated and analysed in a control-focused assessment, the assessor must make an implicit assessment based on expert intuition to determine the risk of a control not being implemented. In a threat-focused approach, each control is traceable to the threat which caused it to be derived, simplifying the analysis of the impact of not implementing a relevant control on the residual cybersecurity risk to the system.

Ultimately, within the context of this research, a threat-focused assessment approach is one which places primacy on the cybersecurity threats to the system, with the assessed vulnerabilities and implemented controls being derived from, and traceable to, those threats.

1.3.4 Model-Based

Model-based analysis techniques are a key focus of this research. This is referring to the tools and techniques that are used within model-based engineering (MBE) and model-based systems engineering (MBSE) more specifically. These model-based approaches are usually defined in contrast to the traditional document-based approaches used to describe a system design.

A document-based approach relies on primarily textual documentation to capture design information, and is defined as:

> *An information system may be said to be a document-based system when it is based primarily upon a store or collection of documents, rather than a store or collection of structured data. [36]*

While a model in the context of a model-based design approach is indeed a collection of structured data, it is more than just a repository of that structured data. A system model is a digital abstraction of a real or to-be-realised system and is characterised by the ability to provide a consistent set of structured information that can be queried by stakeholders to present different views of that same consistent information [37].

Using that understanding of a system model, model-based engineering (MBE) is defined as:

> *Engineering practices in which models are the central and indispensable artifacts throughout a product's lifecycle encompassing concept, development, deployment, operation, and maintenance. [38]*

Model-based system engineering (MBSE) is a form of model-based engineering that focuses on the process of systems engineering, and is defined by the International Council of Systems Engineering (INCOSE) as:

> *the formalized application of modeling to support system requirements, design, analysis, verification, and validation activities beginning in the conceptual design phase and continuing throughout development and later life cycle phases. [39]*

To support model-based system engineering efforts, numerous tools and techniques have been developed and refined by both academia and industry and are available to systems engineers to create a system model. This has helped systems engineering artefacts to transition away from documents towards integrated digital models.

The contemporary approaches to cybersecurity assessment discussed in Section 1.2.1 are primarily document-based, with the results of the assessment captured in spreadsheets and/or text-based documents. Thus, within the context of this research, leveraging model-based analysis techniques means capturing the threat-focused assessment described in Section 1.3.3 using a systems engineering model. This transition of cybersecurity assessments from a

document-based system to a model-based system would mimic the evolution of engineering design from a document-based system to a model-based one, and a research goal would therefore be an investigation of the effectiveness of that potential transition.

1.3.5 Informed Decision-Making

Informed decision making is a key component of this research. In that context, an informed decision is one where the decision maker understands the analysis that has resulted in a recommendation to accept or reject a residual risk level and can apply their own intuition and expertise to review that analysis and decide independently whether to accept the residual risk. An example of an uninformed decision making process would be where the cybersecurity analysis underpinning a recommendation to accept or reject a residual risk was opaque or unintelligible to the decision maker, forcing the decision maker to make a decision based on how much they trust the people making the recommendations to accept or reject a risk because they do not understand or cannot sensibly follow the thought process that led to the recommendation.

To expand on this definition, a generic process for cybersecurity decision making in a risk-based paradigm is shown in Figure 1-3. The process begins with a design activity, performed by the system designer, that incorporates a level of cybersecurity functionality (step 1). A security audit is then performed by cybersecurity subject matter experts (step 2) where the system design is audited against the selected control baseline. As articulated in Section 1.3.3, this control

baseline could be derived from a threat model in a threat-focused approach, or from an existing best-practice control baseline in a control-focused approach. The results of this audit are then used to determine a residual risk (step 3), where the likelihood and consequence of a cybersecurity attack are determined using the results of a holistic security assessment.

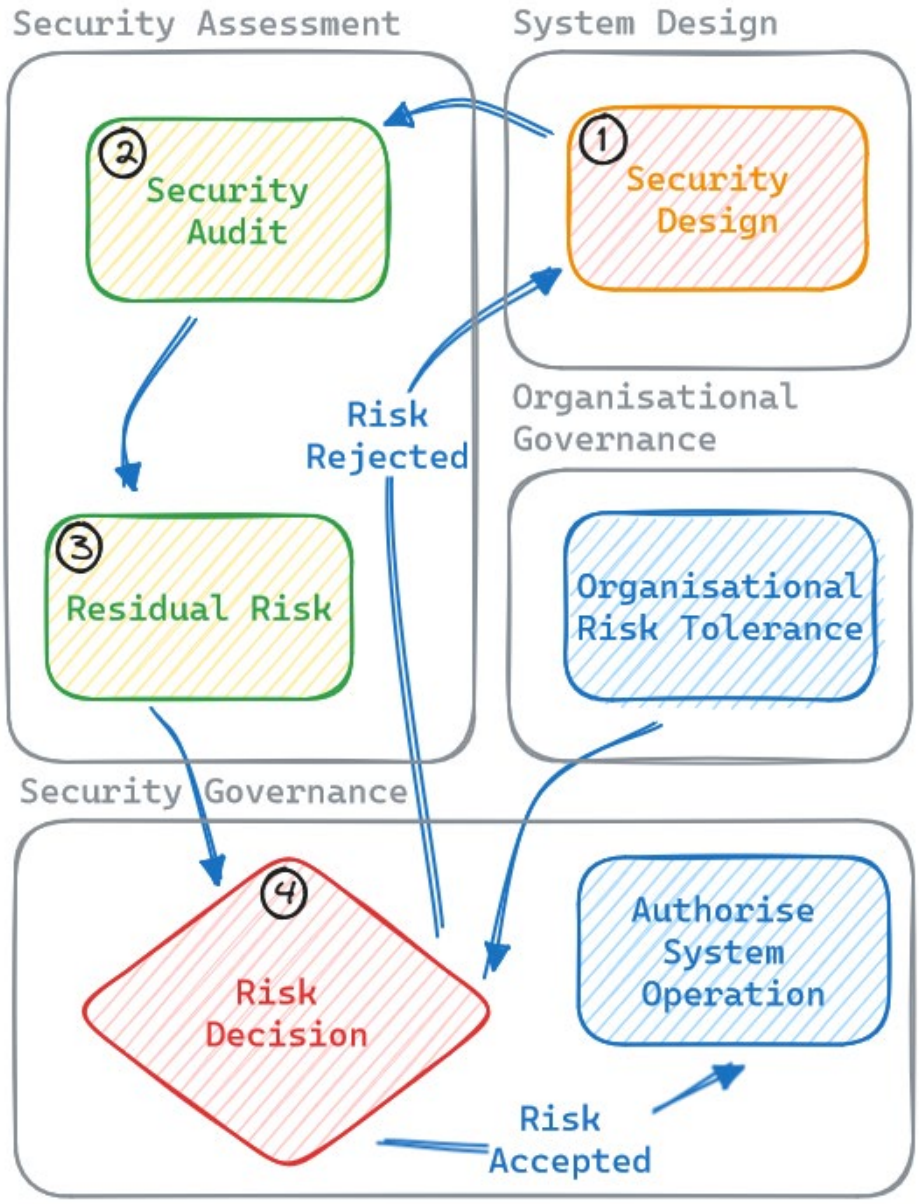


Figure 1-3: *Generic process for making cybersecurity risk decisions.*

This risk assessment is then considered by the security governance functions to make a risk decision (step 4), where the assessed risk is compared to an organisational risk tolerance. If the risk is acceptable, the assessed system is authorised for use. If the risk is rejected, additional mitigations must be implemented with the system design, returning the process to step 1. Obviously, in a realistic situation, the actual process is likely to be more iterative between the first 3 steps, to mitigate against a costly rejection in step 4, but this generic process is broadly indicative of a typical cybersecurity risk assessment process.

For the decision maker to make an informed decision, the reasoning and judgement inherent in the security assessment phase (steps 1 and 2) needs to be understandable to them. As these decision makers are rarely the same people who are performing the assessment, and often do not have the same level of technical cybersecurity expertise as those performing the analysis, the assessment requires the following properties to ensure an informed decision can be made:

1. Intuitive – the assessment must be documented in a manner that is consumable by a non-cybersecurity expert;
2. Traceable – the different factors and variables within the assessment must be explicitly linked to each other, to ensure the contribution of each factor is apparent; and
3. Transparent – any factors, assumptions, judgments and input variables must be clearly identified and justified.

When these properties are adequately demonstrated, the decision maker can understand the assessment and contribute their own expertise to the problem, supplemented by the recommendations of the cybersecurity experts. When the assessment is unintuitive, disparate and opaque, the decision maker is not given the opportunity to understand the analysis and the reasoning underpinning the expert recommendation. They must blindly accept or reject the risk relying entirely on the recommendation of the assessor. This ability to understand the assessment rather than simply accepting it is the defining factor for an informed decision-making process within the context of the research goal.

Of course, one way to ensure this informed decision-making as defined above would be to allow the cybersecurity experts that complete the assessment to make the risk decision (step 4), as they are already fully informed about the assessment. However, this would have the undesirable effect of removing any independence between the security assessment and security governance functions and would likely lead to the decision maker not being adequately informed in terms of organisational understanding and risk tolerance. In practice, organisations developing complex systems typically delegate the decision-making function to a senior corporate officer. The Australian Protective Security Policy Framework [13] provides a clear example of this, where this responsibility for accepting cybersecurity risk within Australian Government departments is given to the department Chief Information Security Officer (CISO). For these reasons, this potential approach of giving the assessors the ability to make the

risk decisions, to achieve informed decision-making, is considered out of scope of a research goal.

1.4 RESEARCH GOAL

The definitions outlined in Section 1.3 provide the foundation for a formalised research goal – that is, to investigate the perceived shortcomings of the contemporary approaches to assessing the cybersecurity risk within complex cyber-physical systems and, if required, develop a new approach to improve efficacy of these cybersecurity risk assessments. The surface-level analysis articulated in the motivation for the research does provide some additional direction toward the possible alternatives for achieving this overall goal and can be used to further narrow the explicit focus of the research.

Firstly, recent thinking from stakeholders in the cyber assessment of cyber-physical systems postulate that threat-focused assessments might be able to overcome the perceived shortcomings of contemporary control-focused approaches. This allows for risk assessments that are contextualised within the specific systems being assessed to ensure that any recommended security mitigations are pragmatic and relevant. Plainly, in an adversarial domain where resources are not infinite, the dogmatic application of ineffective security controls or the over-application of effective security controls in just one area, can have a negative impact on the holistic security posture of a complex system if other critical threat vectors are left vulnerable because of an overinvestment in any single potential vulnerability.

Secondly, the surface-level analysis identified model-based approaches as a potential solution for reusability of threat-based approaches. Modular reusability could help improve the ongoing efficiency of risk assessments across a large organisation, potentially allowing threat-based approaches to achieve similar efficiencies to contemporary control-based approaches. The importance of timely assessment was discussed, making the efficiency of the overall process and the change management process a critical attribute of any risk assessment methodology designed for the constantly evolving cybersecurity environment.

Thirdly, the motivation makes it clear that in an adversarial field without perfect information there is an inevitable reliance on qualitative risk assessment and risk acceptance. In these situations, the assurance provided by the methodology by which the risk is determined can be equally important to the security controls that are eventually implemented. We need rigorous methods that can provide confidence that the risks have been adequately assessed, because by their very definition the acceptance of a risk is acknowledging the possibility of failure and thus the occurrence of an undesirable event cannot be used as a proxy for an adequate assessment process. The earlier sections of this chapter identified that the concepts of worthiness – airworthiness, seaworthiness etc. – that are traditionally applied to materiel and personnel safety assessment could help to demonstrate this assurance in the cyber domain through an equivalent cyberworthiness analysis.

Finally, informed decision-making with respect to cyber risk is another critical component identified in the motivation for this research. Any risk assessment process that requires a balancing of a technical cybersecurity risk against organisational goals and financial realities is dependent on the decision maker truly understanding the risk associated as well as the organisational direction. To allow these decision makers to make informed decisions, a cybersecurity risk assessment must be presented in a manner that can be understood by someone who may not be a technical cybersecurity expert themselves.

The explicit goal of the research is therefore to investigate threat-focused approaches that leverage model-based analysis techniques to assess the cyberworthiness of cyber-physical systems, with a particular focus on the ability of those approaches to facilitate informed decision-making with respect to cybersecurity risk acceptance.

1.5 STRUCTURE

The remainder of this dissertation is structured into the following chapters:

- **Chapter 2 - Literature Review**: Explores the existing literature with respect to the research goal, concluding in an identified gap within the literature that will be addressed by the research captured in this dissertation.
- **Chapter 3 - Methodology**: Outlines the focus of the research, along with a set of formalised research questions and hypotheses. The research

methods used to investigate those research questions will be described, as well as an overview of the expected contribution toward filling the identified literature gaps.

- **Chapter 4 - CEMT Overview**: Describes the Cyber Evaluation and Management Toolkit (CEMT) which was developed to address the research goal, and articulates the ways in which the toolkit can be applied to a cybersecurity evaluation of a cyber-physical system.
- **Chapter 5 - Face Validity Trials**: Includes the results of the face validity trials of the CEMT. This is an embedded chapter that has been published in the MDPI Systems journal.
- **Chapter 6 - Implementation Trials**: Includes the results of the implementation trials of the CEMT. This is an embedded chapter that has been submitted for publication in the Computers and Security journal.
- **Chapter 7 - Conclusion:** Provides a conclusion to the dissertation by discussing the significance of the research captured in the dissertation, outlining how the work has addressed the identified research gap, and describing limitations and avenues for further research in the future.

CHAPTER 2 LITERATURE REVIEW

2.1 INTRODUCTION

To begin our journey toward the research goal described in Section 1.4, we first must understand the current state of the literature to identify whether the literature already fully addresses the motivation behind the research and, if not, whether there are partial solutions that can be leveraged as part of our own contributions back to the literature. This chapter summarises that exploration of the literature and ultimately identifies the gap in the literature that this research project aims to fill.

The literature review focused on three primary concepts within the research goal, namely:

1. Cyberworthiness
2. Threat-focused
3. Model-based

While Chapter 1 also highlighted the terms cyber-physical and informed decision-making from the research goal, these two were not rigorously explored in isolation through the literature review as these terms were primarily providing context for the exploration of the other terms. Cyberworthiness, threat-focused and model-based were all explored within the context of cyber-physical systems as the scope and informed decision-making as the desired outcome.

Through the course of the literature review, the threat-focused term was further subdivided into graph-based cyber risk assessments and mission-based cyber risk assessments, as these were the terms predominately used in the academic literature and professional practice, respectively. Consequently, this chapter will explore relevant literature using the following structure:

- Cyberworthiness; to understand the existing literature with respect to cyberworthiness and similar concepts such as cyber resilience and cyber survivability.

- Graph-based Cyber Risk Assessments – to articulate how graphical approaches to threat modelling have been developed within academic literature.
- Mission-based Cyber Risk Assessments – to describe how professional practice has created threat-focused methods to extend contemporary control-focused approaches.
- Model-based Cyber Risk Assessments – to explore the existing approaches to apply model-based systems engineering techniques to cyber risk assessment.
- Literature Gaps – to summarise the output of the literature review and identify the gaps within the existing literature that this research project aims to fill.

2.2 CYBERWORTHINESS

2.2.1 Background

Cyberworthiness first appeared in the literature when it was introduced by Trope [40] where it was discussed in the context of warranties for software products, provided by their developers, that would provide assurance that the software would be resistant to security breaches. In this paper, Trope provided an analogy to seaworthiness, invoking the assurance and surety that is provided to passengers by the implied warranty of a determination that a vessel is seaworthy, and arguing that cyberworthiness would provide similar assurance and surety over software products in cyberspace.

While the term did not gain traction in academic literature, it was resurrected over 10 years later when members of the Royal Australian Navy discussed the need to consider cyberworthiness when claiming a naval capability is seaworthy [41]. A few years later it was discussed by van der Linden [42], once again in the context of software warranties. Two months later the term cyberworthiness was mentioned in a budget hearing within the Australian Senate [33], with senior officers in the Australian Defence Force (ADF) identifying cyberworthiness as a governance approach that was being taken by the ADF to provide assurance that their systems could operate in a contested cyber environment. The follow-on publications to this discussion in the Australian Senate is where we find the cyberworthiness definition provided earlier in Section 1.3.2:

> *"The desired outcome of a range of policy and assurance activities that allow the operation of Defence platforms, systems and networks in a contested cyber environment. It is a pragmatic, outcome-focused approach designed to ensure all Defence capabilities are fit-for-purpose against cyber threats." [35]*

The literature does not contain much beyond this definition, with no published discussion or proposal for how cyberworthiness could be achieved, let alone be measured or evaluated. This, on its own, represents a significant gap in the literature, but the literature review was expanded to cover two related areas to identify whether the existing literature in these domains could be directly applied to cyberworthiness.

These two domains were:

- Seaworthiness
- Cyber Resilience

In the following sections, the literature within these domains will be discussed in the context of cyberworthiness.

2.2.2 Seaworthiness

There is a significant body of knowledge, both legally and professionally, around the worthiness concepts from which cyberworthiness was derived – particularly seaworthiness. Interestingly, despite its common usage in the maritime domain there is not a single accepted definition for what makes a vessel seaworthy [43], but the most prevalent legal definitions stretching as far back as 1839 [44, 45, 46] define it in terms of the outcome that a vessel is sufficiently fit to undertake a voyage, rather than a set of requirements defining how that would be achieved.

These definitions in isolation are largely unsatisfying from the perspective of trying to predict seaworthiness – or cyberworthiness – as a system designer, as it can imply that a system is worthy until it fails at which point we can conclude that it, by definition, likely never was. This does make it clear, however, that if worthiness is defined by a future outcome, then a worthiness claim is both predictive and uncertain, placing it squarely within the domain of predictive risk assessment.

Helpfully, the framework for assessing seaworthiness within the Australian Defence Force is well established as the Defence Seaworthiness Regulatory System Publication 100 [47]. This framework outlines a methodology for ensuring that the system maximises its ability to perform its specified purpose while minimising safety risk so far as is reasonably practicable (SFARP). Importantly, the framework places a focus on how decisions are made, and describes four key principles for how worthiness decisions should be made:

> *"The seaworthiness governance principles require that seaworthiness decisions are made:*
>
> *a. mindfully – decisions are more effective and less likely to have unintended consequences when they are made with a thorough understanding of the context, the required outcome, the options available, and their implications now and in the future*
>
> *b. collaboratively – obtaining input from all stakeholders and engaging in joint problem-solving results in better decisions (bearing in mind that collaboration does not necessarily require consensus)*
>
> *c. accountably – decisions only become effective when people take accountability for making them happen*
>
> *d. transparently – decisions are more effective when everyone understands what has been decided and why." [47]*

While this provides a set of strong foundational principles that could be applied to cyberworthiness, this seaworthiness management framework also highlights

an inherent link between seaworthiness and safety risk assessments. Given this alignment between seaworthiness and system safety assessment, the literature on the application of safety risk analysis techniques to cybersecurity assessment was also reviewed as this may identify extant methods that could be used for cyberworthiness assessments.

Suo et al [48] provides an example of a merging system safety analysis with cybersecurity analysis, by treating cybersecurity threats as analogous to a safety hazard. This focuses primarily on cybersecurity incidents as a cause to a safety accident, a methodology mirrored in the work of Macher et al [49]. While this approach does integrate the two domains nicely, it does not cover those cybersecurity incidents that do not necessarily lead to a safety accident and would need to be modified to address a loss of functionality due to a realised cyber risk.

More recent efforts [50] have looked to directly apply safety analysis techniques such as Fault Tree Analysis (FTA), Functional Hazard Analysis (FHA) and System Theoretic Process Analysis (STPA) to the cyber domain. The use of a modified version of STPA, referred to as STPA-Sec, was outlined by Friedberg et al [51], though in an analysis of this technique Schmittner et al [52] highlighted that an important difference between safety and security analysis is the intention of a threat actor, and that this must be modelled.

This fundamental difference limits the direct applicability of system safety analysis approaches to cyber risk assessments, but the techniques within those

approaches do assist with providing assurance and could have similar outcomes when adapted to cybersecurity. Graph-based analyses, such as bow-tie diagrams, are examples of those techniques, and they have had some success when adapted to the cyber domain as outlined in Couce-Vieira et al [53] and Couretas [54], primarily because they have been found to provide an intuitive summary of cybersecurity risk.

2.2.3 Cyber Resilience

As outlined earlier, because of the lack of literature related to the specific cyberworthiness term, the literature review was broadened to include closely related terms that are more popular within the literature, namely cyber resilience. This term was considered similar because, like cyberworthiness, it has an explicit focus on the ability of a system to remain functional when under persistent cyber-attack. In doing so, cyber-resilience has a natural bias toward securing the integrity and availability of a system rather than the confidentiality of data stored on that system and has been linked to cyberworthiness within the literature [55, 56].

Cyber resilience as a domain is significantly more mature than cyberworthiness, and there are several international standards and professional best-practices that have been developed to guide the enhancement of cyber resilience [57]. These standards and best practices include traditional cybersecurity standards such as ISO 27001 [58] and NIST 800-53 [19], but also include best-practice guidance more focused on cyber resilience, such as the NIST Cybersecurity Framework

[59], NIST 800-160 Vol 1 – Engineering Trustworthy Secure Systems [60] and NIST 800-160 Vol 2 – Developing Cyber-Resilient Systems [61].

The NIST Cybersecurity Framework (CSF) defines six key functions within its core structure – Govern, Identify, Protect, Detect, Respond and Recover – and these are used to organise cybersecurity outcomes at the highest level. NIST 800-160 Vol 1 provides a process for deriving security requirements in complex systems by assessing the protection needs of the assets within a system and emphasises the importance of the trustworthiness of the assessment and the resultant assurance levels behind that assessment. NIST 800-160 Vol 2 complements the asset-focused methodologies in NIST 800-160 Vol 1 by deriving a set of fourteen cyber resiliency techniques that can be implemented by system designers to improve the cyber resiliency of assets and systems.

There are also methodologies proposed within the literature to evaluate the cyber-resiliency of systems, and they generally fall into two broad categories. The first category is principle-based evaluations, such as the Cyber Resilience Self-Assessment Tool (CR-SAT) [62], which take a governance approach and evaluate systems or organisations against a set of cyber resilience principles, and the second category is dependency-based evaluations [63], where the dependency relationships between assets or systems are modelled using directed graphs to understand resilience through defence-in-depth.

The current state of the literature for the cyber resilience of cyber-physical systems specifically was explored in a comprehensive and systematic literature

review by Segovia-Ferreira et al [64], which concluded that there were three open challenges within that literature:

1. System Modelling: the literature could be improved by methods for higher fidelity attack models and the performance of systems against those attacks
2. Metrics and Evaluation Methods: the current literature doesn't deal with the complexities of holistic interactions between systems and controls when evaluating the overall resilience of a system-of-systems
3. Testing and Validation Environments: the validation of approaches within the literature primarily use constrained testbeds and there is a lack of scalability validation that validates the approaches when applied to complex real-world systems

2.2.4 Review

The literature on cyberworthiness is largely immature and this creates a general literature gap with respect to cyberworthiness, particularly around how one might demonstrate or evaluate cyberworthiness. The literature in the related domains of seaworthiness, system safety and cyber resilience do provide some promising foundations for exploring cyber worthiness, particularly the seaworthiness governance principles, the graph-based approached used in both system safety and cyber resilience, and the best-practice cyber resilience principles defined within the relevant NIST standards. However, the absence of an adversarial actor in system safety approaches limits their direct applicability to cyberworthiness

and gaps currently remain within the cyber resilience literature related to evaluation, detailed modelling and large-scale validation.

Figure 2-1 outlines the relationship between seaworthiness, cyber resilience and cyberworthiness, as well as the cyberworthiness principles derived from the seaworthiness literature. These principles formed the target of subsequent literature review efforts, looking through existing threat-focused cyber assessment techniques to understand the extent to which they can address these cyberworthiness principles.

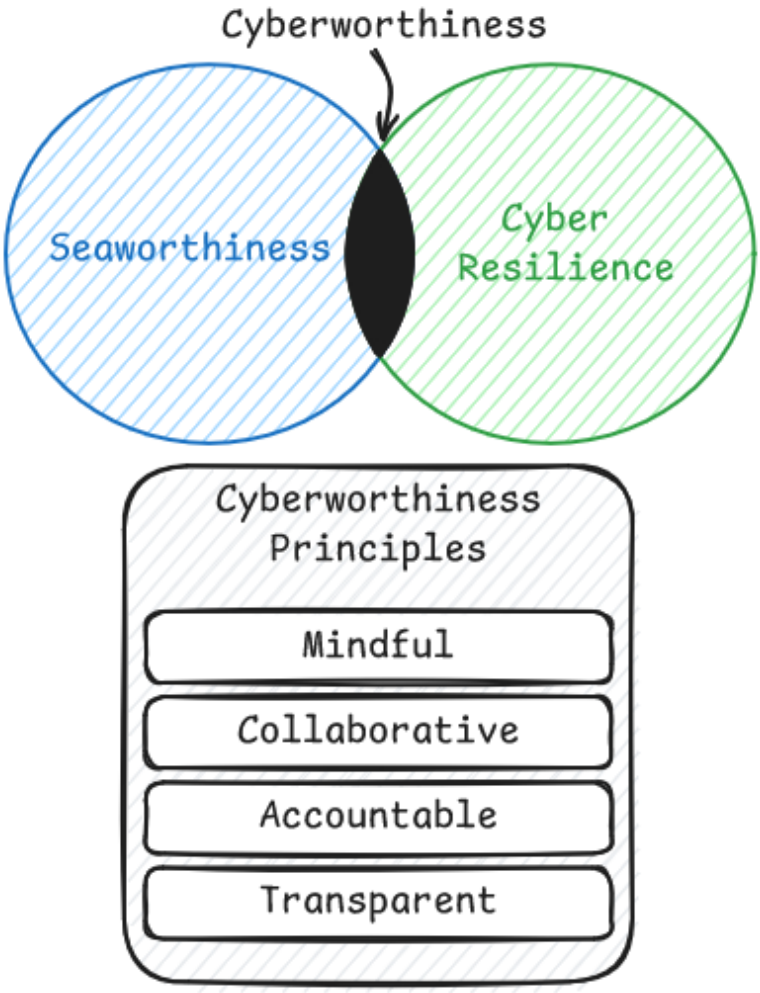


Figure 2-1: *Cyberworthiness principles.*

2.3 GRAPH-BASED CYBER RISK ASSESSMENTS

2.3.1 Background

Graph-based Cyber Risk Assessments is the first type of threat-focused assessment methodology explored through this literature review. Graph-based techniques were identified during the literature review of both system safety and cyber resilience as techniques that had been leveraged to complete analysis and assessment in those domains. Graph-based techniques are also prevalent in foundational threat modelling approaches used within professional cybersecurity practice [65]. For these reasons, graph-based cyber risk assessments became a key focus of this literature review.

Graph-based cybersecurity analysis is a research area that focuses on the development of graphical models to analyse cybersecurity scenarios. According to Kordy et al:

> *"the great advantage of graph-based approaches lies in combining user friendly, intuitive, visual features with formal semantics and algorithms that allow for qualitative and quantitative analysis." [66]*

The intuitive nature of these approaches shows promise for facilitating collaboration with stakeholders who may not have a deep cybersecurity background. Kordy et al [66] categorised graph-based security analyses into two broad categories – statically modelled graphs and sequentially modelled graphs – which can in turn each deal with either only attacks or both attack and defence.

This literature review focuses in on two types of graph-based security analyses. Firstly, attack trees, which are traditionally statically modelled graphs but have been extended and reapplied as sequentially modelled graphs more recently, and misuse case graphs, which are primarily sequentially modelled graphs.

2.3.2 Attack Trees

Attack trees, first outlined in Weiss [67] and popularised by Schneier [68], are the basis for many of the statically modelled graphs used for cybersecurity analysis. These graphs depict a hierarchical tree structure which describes the combination of events that need to occur for a cybersecurity risk to be realised and are broadly similar to the fault trees common to system safety analysis. Attacks trees have been successfully incorporated into the likelihood calculations of traditional risk assessment methodologies [69] to support graph-based cyber risk assessments. Attack trees have also been extended to include the depiction of defensive actions in Kordy et al [70] and termed attack-defence trees.

These attack trees traditionally fall into the category of statically modelled graphs but have also been adapted and enhanced within more contemporary literature to function like a sequentially modelled graph. The sequentially modelled graphs outlined in the literature are comprised of a sequence of events that an attacker must perform to compromise a target system.

The literature contains several variants of these sequential attack trees, from serial attack trees [71] to attack graphs [72, 73, 74] and attack chain tree [75].

More recently, the risk likelihood assessments that flow from sequential attack trees have been estimated by converting the attack tree structure into a Bayesian network to combine probabilities of the various nodes on the attack tree [76, 77]. This primary benefit of these sequentially modelled attack trees is that they provide a focus on the intent of an attacker; one of the key attributes of a threat-focused approach to cyber risk assessments.

2.3.3 Misuse Case Graphs

While attack trees began as statically modelled graphs and over time transitioned toward sequentially modelled graphs, there have been other types of graph-based cybersecurity analysis techniques that have always been sequentially modelled. The concept of insecurity flow graphs was introduced by Moskowitz and Kang [78], to model an end-to-end attack sequence, though this concept attempted to reduce the graph into a mathematical model, losing some of the intuitiveness benefits of a graph-based approach in the process.

The construction of misuse case graphs, also known as abuse case graphs, is another approach in the literature that focuses on the actions of the attacker and their interaction with the target system. Appearing in the literature at the turn of the century in McDermott [79] and Sindre et al [80], misuse cases borrow from the use case concept common in software development practices but focus on how an adversary, rather than an authorised user or operator, interacts with a system. More recent extensions to the misuse case concept include overlaying the misuse cases with the system architecture [81] and developing misuse case

coverage metrics as a proxy for security risk [82], but on their own misuse cases require extensive textual encoding to adequately convey the full risk picture [83].

While misuse cases are often used to elicit security requirements [84], mirroring the purpose of use cases in software development, Matuleviaius [85] has linked misuse cases to the mal-activity graphs proposed by Sindre [86]. These mal-activity graphs provide a sequential model of the activities taken by the attacker – much like the insecurity flow graphs introduced by Moskowitz and Kang [78] – and can be used to provide additional detail to support misuse case graphs.

2.3.4 Review

The literature for graph-based cyber risk assessments is relatively mature, with a well-defined set of techniques that have been widely applied and investigated. Within the context of the research goal, there are no gaps within graph-based cyber risk assessments specifically that will be addressed by this research. However, both attack trees and misuse case graphs have been identified as existing techniques within the literature that could be utilised to help address the research goal.

The literature does include comparison studies between the attack tree methodologies and the misuse case methodologies. Both Opdahl et al [87] and Karpati et al [88] found that while attack tree methodologies tended to identify more threats, there were significant differences between the types of threats found by each technique. This would suggest that a combined approach, such as

that outlined by Tondel et al [89], which looks at both misuse cases and attack trees might be the most effective approach.

Of course, the efficiency in risk assessment is also an important concern, and while there has been some literature looking to increase the reusability of these graph-based threat models [90] these techniques still require knowledgeable experts to apply them to systems individually with limited reusability within or across assessments.

Based on this literature review, the primary benefit expected from incorporating these graph-based cyber risk assessment techniques into a threat-focused approach to a cyberworthiness evaluation is the ability to perform detailed adversary modelling in a manner that is both intuitive and understandable. As depicted in Figure 2-2, this intuitiveness of graph-based techniques can address the transparent and accountable cyberworthiness principles.

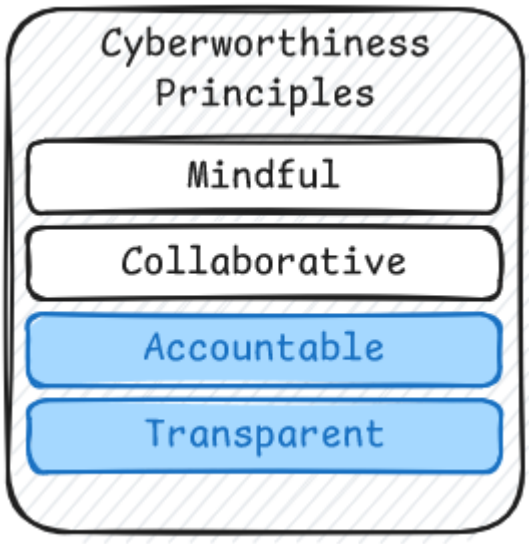


Figure 2-2: *Cyberworthiness principles addressed by graph-based cyber risk assessments.*

2.4 MISSION-BASED CYBER RISK ASSESSMENTS

2.4.1 Background

Mission-based Cyber Risk Assessments (MBCRA) were identified during the literature review as a domain that warranted further investigation because professional cybersecurity practitioners were using this technique to test and evaluate cyber resilience and consequently, MBCRAs were uncovered during the earlier exploration of the cyber resilience literature. MBCRAs are required by the US Department of Defense (DoD) as part of its Cybersecurity Test and Evaluation (T&E) procedures [91] and because of this MBCRAs have been widely adopted by professional cybersecurity consultancies that work with the US DoD [92, 93, 94] and also by groups within the US DoD itself [95, 96].

Figure 2-3 depicts the critical elements in an MBCRA, starting from a set of input artifacts and leading to a set of outputs that summarise the cybersecurity analysis that was performed and provides appropriate recommendations. Importantly, there is a strong threat-focus throughout the process, starting from initial threat assessments and attack surface characterisations during the *Input* phase, collaborative analysis of attack paths in a mission context during the *Execution* phase and reporting on specific cyber-attack scenarios during the *Output* phase.

The explicit focus on collaboration and contextualisation to the purpose and functionality of the system being evaluated also provide alignment with the cyberworthiness principles derived in Section 2.2.4.

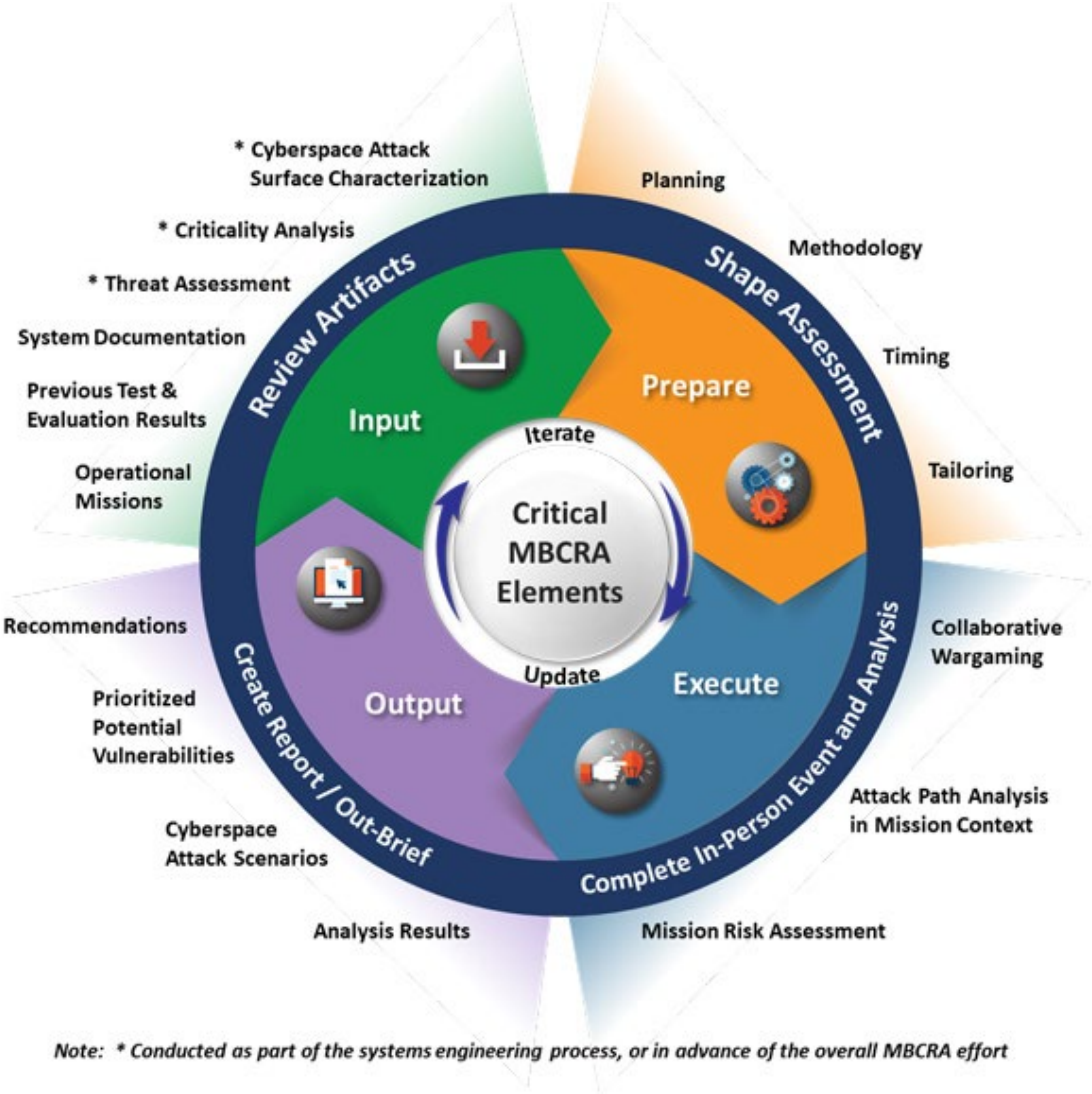


Figure 2-3: *Critical MBCRA elements, reproduced from [97].*

In terms of a formal definition of an MBCRA, according to the US DoD Cybersecurity Test and Evaluation Guidebook:

> *"MBCRA is a process for identifying, estimating, assessing, and prioritizing risks based on impacts to DoD operational missions resulting from cyber effects on the system(s) employed." [98]*

However, this is perhaps a vague definition of the MBCRA itself, as within the context of the broader cybersecurity test and evaluation framework defined by the US DoD, the MBCRA is essentially a threat-focused assessment that is designed to complement the RMF, which is one of the contemporary control-focused approaches discussed earlier in Section 1.2.1.

While there are many differing methodologies that can be used to meet the requirement for an MBCRA [98], the most prevalent methods for doing this are a Cyber Table Top (CTT) and the Mission-based Risk Assessment Process for Cyber (MRAP-C) [99], and these two methods will be the focus of further exploration of MBCRAs for this literature review.

2.4.2 Cyber Table Top (CTT)

The Cyber Table Top (CTT) methodology is comprehensively described in the US DoD Cyber Table Top Guide [100], and this guide summarises the CTT in the following way:

> *"The Cyber Table Top (CTT) is a focused, intellectually intensive exercise that explores the effects of cyber offensive operations on the capability of US systems to carry out their missions. It is a wargame-like exercise that centers on two teams with opposing missions: the military forces charged with executing an operational mission and the cyber mission forces attempting to oppose those military forces." [100]*

These two teams are referred to as the Operational team, representing the system operators and defenders, and the Opposing Force (OPFOR) team, representing the cyber force trying to compromise the system and its mission. These teams are composed of stakeholders in the design and operation of the system, including engineers, operators, designers, program personnel and cybersecurity experts. Importantly, the process is non-adversarial, and the goal of the process remains the characterisation of the system's cybersecurity rather than one team or the other 'winning' by overcoming the other team. [100]

Due to this non-adversarial approach, the Cyber Table Top is characterised as a collaborative MCRA methodology [99]. The primary outputs generated by the process are risk matrices that summarise the risks to the system, based on the expected effect those risks have on the system's ability to perform its mission, and a set of recommended actions the could improve the system's resilience to cyber-attacks [100].

Analysis of practical applications of CTTs found that other outputs of the CTT process were threat scenarios, which are defined as "*specific threats or conditions associated with the impact or consequence to the system in that circumstance*" [99], and a set of key interfaces that required further testing. The threat scenarios were seen by practitioners as particularly important as a communication tool to inform decision makers because they illustrated the results of the analysis in an understandable manner [99].

2.4.3 Mission-based Risk Assessment Process for Cyber (MRAP-C)

The Mission-based Risk Assessment Process for Cyber (MRAP-C) methodology was developed by the US Air Force to provide a framework for an integrated and iterative Mission-based Cyber Risk Assessment (MBCRA). It incorporates best practices from the CTT methodology, and extends those processes to create an approach that uses systematic risk analysis to characterise failure modes within a system and identify potential cyber vulnerabilities [101].

The MRAP-C includes a detailed functional thread analysis to understand and characterise the system's attack surface which is then mapped to system functions and potential cyber vulnerabilities. Risk ratings and priority levels are assigned to those potential cyber vulnerabilities and vignettes describing attack paths that could exploit those vulnerabilities in a manner that would have a mission impact are generated [101].

Similar to the CTT, the MRAP-C is a collaborative MBCRA methodology [99], that brings together an integrated team of designers, operators and cybersecurity experts to explore and evaluate the cybersecurity risks associated with a system. Ultimately, the goal of the MRAP-C is to find vulnerabilities that could cause a mission failure due to a loss of system functionality [101].

Analysis of the practical applications of MRAP-C found that it included all the outputs of the CTT but also added a prioritised list of mitigations that could be applied to the system as well as a set of attack graphs or attack [99].

2.4.4 Review

Mission-based Cyber Risk Assessments (MBCRA) provide examples from professional cybersecurity practice of methodologies that use threat-focused approaches to complement the contemporary control-focused methodologies to address cyber resilience concerns. These assessment methodologies provided a specific focus on the mission-critical functions performed by the system, ensuring that the outcome of the analysis was relevant to the holistic context of the system.

The MBCRA methodologies documented the assessments using threat scenarios and vignettes to explain the risk to decision makers, and the MRAP-C process also incorporated attack graphs and attack trees to provide further detail behind these vignettes. These output formats were found to be useful tools to communicate risk information to stakeholder who may not be experts in cybersecurity.

The MBCRA literature showed that there are existing approaches within professional cybersecurity practice that take a threat-focused approach to cyber resilience assessment. In terms of addressing the cyberworthiness principles, it is expected that MBCRAs would be able to address the collaboration and mindfulness principles but would likely need extension or modification to address the transparency and accountability principles.

The attack graphs and vignettes used by the MBCRA methodologies provide a good summary of the attack paths, but a more detailed assessment that leverages other graph-based techniques would be able to provide additional detail that would support decision maker accountability. Additionally, the qualitative risk analysis used within MBCRAs may not be able to provide the necessary transparency in how those qualitative decisions are made to meet the intent of the related cyberworthiness principle. Generally, it is not clear what level of assurance is demonstrated through the MBCRA processes as it relies on the participants collaborating effectively and consistency across different activities must be achieved manually.

As displayed in Figure 2-4, the mission-focus and explicit collaboration within MBCRA processes can address the two cyberworthiness principles that were not addressed by graph-based cyber risk assessment techniques – mindful and collaborative.

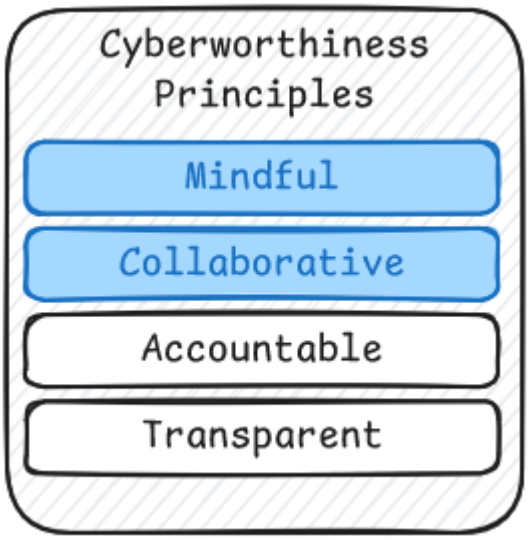


Figure 2-4: Cyberworthiness principles addressed by mission-based cyber risk assessments.

Overall, the professional literature around MBCRAs is quite mature, even if the academic literature is not as developed. However, many of the tools that support MBCRAs are not openly available outside of the US military, and this limits their ability to be leveraged within academic literature. Nonetheless, it is apparent from the literature review that the approaches that are taken within MBCRAs, if enhanced with additional graph-based techniques, would provide a strong foundation for addressing the derived cyberworthiness principles.

2.5 MODEL-BASED CYBER RISK ASSESSMENTS

2.5.1 Background

While the earlier sections of this literature review showed that graph-based cyber risk assessment techniques and mission-based cyber risk assessment processes could be combined to address the derived cyberworthiness principles, an overall methodological approach is required to unify those assessments in an effective and efficient manner. As briefly explored in Section 1.3.4, model-based approaches present a promising potential solution for capturing a threat-focused cyber risk assessment in a way that supports traceability, reusability and consistency.

Model-based Cyber Risk Assessments is the term given to processes that leverage existing model-based systems engineering tools and techniques to create cybersecurity risk assessments. The literature includes model-based approaches to cybersecurity assessment which involve the development of a

meta-model or consistent language to aid in the scalability and reusability of threat modelling activities. Generally, the application of security analyses, including graph-based security analyses, can be time consuming and "model-based security assessment methods can potentially be used to streamline this process" [102]. Additionally, these model-driven approaches often include the capability of being incorporated into a system-level computational simulation, which can support both quantitative and qualitative methods, allow for reusability as the threat profile changes over time and can offset the reliance on subjective expert opinion for assessing risk [54]. Model-based approaches to system design and development have also been shown to increase transparency through explicit traceability between decisions and the data justifying those decisions [25], and this is likely to carry across to cyber risk assessments performed using model-based tools.

To assess the applicability of existing model-based cyber risk assessment approaches within the literature, a set of expected attributes that would contribute to an effective model-based cyber risk assessment was defined. From these potential benefits of model-based cyber risk assessments, the following attributes were derived:

- Extended model-based taxonomy – Inheritance of existing model-based functionality through extension of model-based systems engineering languages to support the reusability and maintainability of threat models

- Visualisation & simulation of threats – threat modelling is expressed using intuitive visualisations and risks can be assessed through computational simulations
- Explicit traceability to threats – derived security mitigations are traceable to adversary actions to facilitate review of the importance of those mitigations

In addition to these three attributes that were derived from the potential benefits of using model-based techniques themselves, two more attributes were defined to further incorporate the graph-based and mission-based cyber risk assessment techniques discussed earlier in this literature review:

- Threat focused – to incorporate the benefits of mission-based cyber risk assessments, the threats to the system are used as the focal point of the assessment
- Detailed adversary modelling – to incorporate the benefits of graph-based cyber risk assessments, the actions of the adversary are modelling in detail, facilitating a precise discussion and review of the resulting threat analysis

These five attributes will be used to determine the suitability of existing model-based cyber risk assessment methodologies to address the defined research goal. In the following sections, the extant methodologies for model-based cyber risk assessments will be discussed, followed by an analysis of the suitability of these methods against the five defined attributes.

2.5.2 Existing Methodologies

In 2015, Nguyen et al. [103] provided a comprehensive review of the literature associated with using model-based techniques for the security assessment of cyber-physical systems. This state of research was revisited by Geissman and Bodden [104] in 2020. They found several approaches and methodologies in the literature that leveraged model-based techniques, including SysML-based approaches, to address the security assessment of complex cyber-physical systems.

Many of these existing approaches, such as the Cyber Security Requirements Methodology (CSRM) [105] and Larsen et al. [106], are focused on the definition of security requirements and provide a methodology for managing the implementation of those security requirements through the systems design lifecycle. These approaches are not explicitly threat focused and are primarily concerned with the management of existing cybersecurity requirements throughout the design and production of a system, rather than deriving those requirements from a threat model.

Several other methodologies, including SAVE [107], Surreal [108], Navas et al. [109] and Geissman et al. [110], are also threat focused and utilise graph-based threat modelling techniques such as misuse case graphs. However, these approaches provide only a high-level threat model that interacts with a detailed white-box model of the system. Without extending those misuse case graphs into more detailed graphs and models of adversary behaviour against the system,

these approaches are limited in their ability to facilitate precise discussions of threat activities when reviewing the threat analysis.

Some emerging methodologies, such as MBSESec [111], also provide more detailed modelling of adversary behaviour in the form of activity diagrams. However, these approaches do not include explicit control traceability to the detailed threat model and the use of specific SysML profiles that support the rapid creation of summary visualisations such as attack trees. While this traceability and visualisation may not necessarily produce a better or more accurate threat model, it would enhance the ability of non-specialists to interact with and critically review the threat model to make more informed risk decisions.

UAF [112], UMLSec [113] and SysML-Sec [114] all provide taxonomies and extension profiles that focus on the allocation of security requirements and the linkages of assets, vulnerabilities and mitigations. These approaches are not explicitly threat focused and instead focus on the allocation of existing security controls and requirements to assets within a system, rather than deriving those requirements from the threats to the system. Without those threats being a focus of these approaches, they are limited in their ability to demonstrate traceability to threats or simulation of those threats.

2.5.3 Review

Table 2-1 provides a comparison of the published model-based cyber risk assessment methodologies against the five key attributes described in Section

2.5.1. While all the methodologies addressed the first attribute by providing an extended taxonomy, no methodology addressed all five attributes. None of the reviewed methodologies adequately incorporated computational simulations into their threat assessment methodologies, nor did they provide explicit traceability between detailed threat models and a resultant set of mitigations.

Table 2-1: *Comparison of Model-based Cyber Risk Assessment methodologies*

	Model-based Security Assessment Approach	Extended Model-based Taxonomy	Threat Focused	Detailed Adversary Modelling	Visualisation & Simulation of Threats	Explicit Traceability to Threats
1	CSRM [105]	Y	N	N	N	N
2	Larsen et al. [106]	Y	N	N	N	N
3	SAVE [107]	Y	Y	N	N	N
4	Surreal [108]	Y	Y	N	N	N
5	Navas et al. [109]	Y	Y	N	N	N
6	Geissman et al. [110]	Y	Y	N	N	N
7	MBSESec [111]	Y	Y	Y	N	N
8	UAF [112]	Y	N	N	N	N
9	UMLSec [113]	Y	N	N	N	N
10	SysML-Sec [114]	Y	N	N	N	N

Overall, the maturity of the literature with respect to model-based cyber risk assessments is mixed. There are several existing methodologies that use model-based techniques to assess the cybersecurity risk of cyber-physical systems, but they do not reach the full potential of model-based approaches to cybersecurity analysis. Figure 2-5 provides a visual depiction of the level to which the five key attributes identified in this literature review are served by existing model-based cyber risk assessment methodologies, with green representing those attributes that are well served by existing techniques, yellow indicating attributes that are served by some of the existing techniques, and red indicating those attributes that are not currently found in existing techniques.

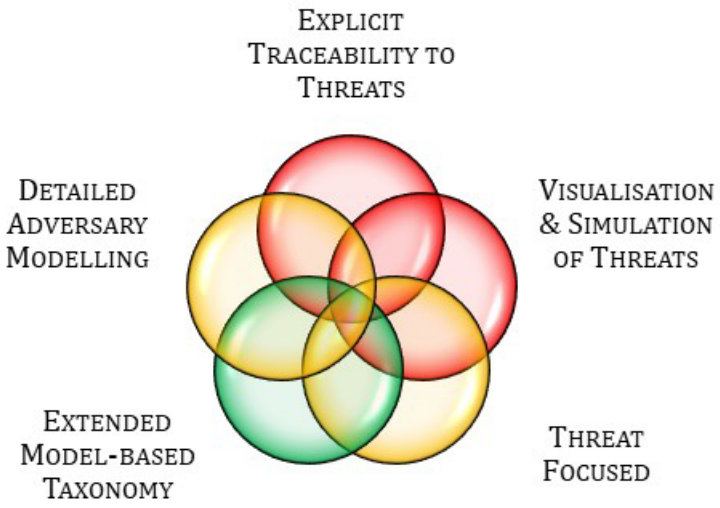


Figure 2-5: *Summarised maturity of the Model-based Cyber Risk Assessment literature*

This represents a significant gap in the literature for model-based approaches to threat-focused cyberworthiness assessments of cyber-physical systems. To address this gap, a model-based cyber risk assessment methodology would need to be developed that builds upon the existing state-of-the-art by incorporating explicit traceability between detailed threat models and derived security mitigations and provide visualisation and simulation of threats.

2.6 LITERATURE GAPS

This literature review investigated four domains to explore the ability of the existing literature to satisfy the motivation and research goals articulated in Chapter 1. These four areas were:

1. Cyberworthiness
2. Graph-based Cyber Risk Assessments

3. Mission-based Cyber Risk Assessments
4. Model-based Cyber Risk Assessments

The literature associated with cyberworthiness was found to be rather immature, with little specific investigation into the concept of cyberworthiness or how it might be achieved. Related work in the fields of seaworthiness and cyber resilience identified four foundational principles that could underpin a cyberworthiness claim, on which decisions are made:

- Mindfully
- Collaboratively
- Accountably
- Transparently

The literature on graph-based cyber risk assessments was relatively mature, with several techniques that could be used to provide intuitive and understandable cyber risk assessments. These graph-based techniques could also help to provide detailed threat modelling to support the documentation of adversary actions against a system. However, these graph-based techniques on their own are insufficient for assessing the cyberworthiness of cyber-physical systems without a holistic methodology encapsulating them.

The literature related to mission-based cyber risk assessments was also quite mature, especially in professional cybersecurity practice, but the key limitation of this literature is that implementations of these methodologies are not openly

available, because they are either proprietary or are otherwise controlled in their distribution. The information that is available of the overall mission-based cyber risk assessment approaches do provide a robust framework for ensuring a threat focused assessment that is mindful of the system context and the operational context of the system.

The literature around model-based cyber risk assessments was in a developing state, with several existing methods that used model-based approaches to assess the cybersecurity of a system. However, none of these existing methods adequately address the five attributes of a model-based cyber risk assessment process that would be required to adequately evaluate the cyberworthiness of cyber-physical systems, namely:

- Extended model-based taxonomy
- Threat focused
- Detailed adversary modelling
- Visualisation and simulation of threats
- Explicit traceability to threat

Figure 2-6 summarises the maturity of the literature associated with these four domains, with red indicating a low maturity, orange indicating a moderate maturity and green indicating a high maturity. The literature review has identified that each of these four domains can contribute to the evaluation of cyberworthiness in cyber-physical systems, but the none of them can achieve this on their own.

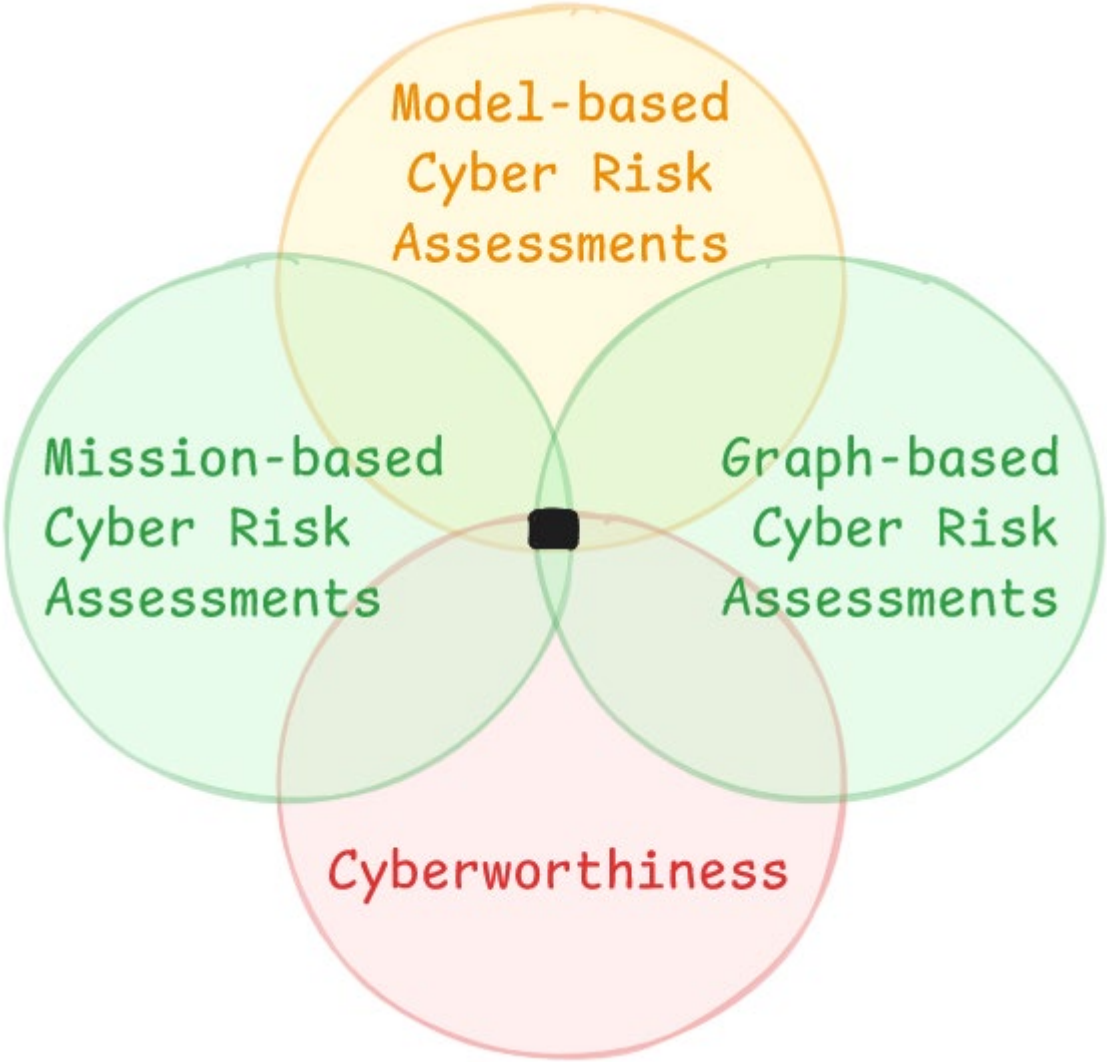


Figure 2-6: *Literature Gap*

Consequently, the literature gap that has been identified is the intersection of these four areas – using mature mission-based and graph-based techniques in a model-based approach to meet the five desired attributes of a model-based assessment and the four foundational principles of cyberworthiness. This provides a focus for the research project and a basis for the significance of the work that has been undertaken.

CHAPTER 3 METHODOLOGY

3.1 RESEARCH FOCUS

The focus of the research captured in this dissertation is to build an understanding of threat-focused approaches to address the literature gaps identified in Chapter 2. The key themes resolve around the ability to understand the cyber risk of cyber-physical systems through rigorous engineering methods and effectively communicating that risk to decision makers.

Given the pedigree of cyberworthiness within the Australian Defence Force (ADF), the focus of the initial trials will be on systems operating within the

regulatory paradigm of the ADF. This focus is both important and topical, as Defence forces around the world operate not only in an adversarial cyber domain but also an adversarial functional domain, where adversaries can move along the spectrum from cooperation to competition to conflict and back again over time. In this environment the ability to make rapid, informed risk decisions backed by rigorous assurance is critical, because the real consequences to national security and human safety of not deploying a critical military capability need to be weighed against the equally real consequences to national security and human safety of operating a critical military capability that is not cyberworthy. This environment where the stakes are so high, but there is no safe side upon which to err, is largely unique to the Defence environment.

Simultaneously, the research must not become over-optimised for the ADF context, and must be generalisable to suit other similar domains, such as commercial cyber-physical systems or critical infrastructure. However, by focusing on the Defence context, the expectation is that if the framework is shown to be feasible within that more rigorous environment it could then be adapted into other similar domains that need to conduct cyber risk assessments of cyber-physical systems.

There will also be a strong focus on practical applications that move beyond theoretical frameworks and implementations on idealised systems, towards trials and investigations that are performed on critical cyber-physical systems under real-world conditions. This is primarily because in highly regulated environments,

it is often difficult to translate a theory into a practical and feasible implementation. Completing that part of the work during this research will provide a greater professional contribution than only developing a framework. These practical applications will also help to demonstrate the relative efficiency of any framework under real-world conditions, which is equally critical to the accuracy and effectiveness of the framework, especially in the adversarial environments discussed earlier.

In the following sections, the specific research gaps will be identified, along with an articulated aim for the overall research effort and a hypothesis for a potential methodology to address that aim.

3.1.1 Research Gaps

In Chapter 2, a literature gap was identified at the confluence of four areas within the existing literature:

- Cyberworthiness
- Graph-based Cyber Risk Assessments
- Mission-based Cyber Risk Assessments
- Model-based Cyber Risk Assessments

The overlap between these areas is a domain where the threat-focus and operational lens of the mission-based assessments are defined using the structured languages and numerical simulation capabilities of model-based assessments to achieve the engineering rigour demanded by cyberworthiness

and is expressed with the intuitiveness and explainability of graph-based assessments. As shown in Figure 3-1, this confluence is termed *Threat-based Cybersecurity Engineering*.

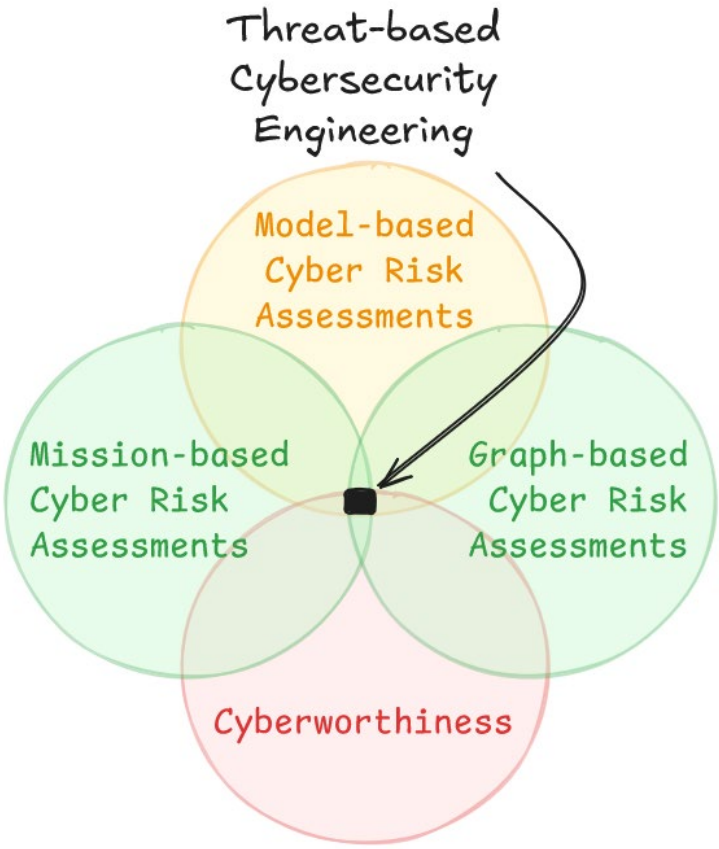


Figure 3-1: *Filling the Literature Gap - Threat-based Cybersecurity Engineering*

However, as discussed earlier, the focus of this research was on applied research outcomes, and so the research gap we were trying to address was not necessarily the lack of a rigorous definition and theoretical framework for Threat-based Cybersecurity Engineering. The research gap we were trying to address was the lack of a practical implementation, backed by applied research that evaluates that implementation against extant contemporary approaches, of a process for undertaking Threat-based Cybersecurity Engineering.

3.1.2 Aim

The explicit aim of this applied research was to identify a set of guiding principles for Threat-based Cybersecurity Engineering and to use those principles to develop an implementable toolkit that can be used by system designers and cyber experts to conduct cyberworthiness evaluations of cyber-physical systems. The ability to implement this toolkit in a real-world system was investigated to assess the practical effectiveness and efficiency of this toolkit at evaluating cyberworthiness to determine whether the toolkit could be implemented in practice.

3.1.3 Hypothesis

The hypothesis was that Threat-based Cybersecurity Engineering would be an effective and efficient approach to the assessment of cyberworthiness in cyber-physical systems. As outlined in Section 3.1.1, Threat-based Cyber Security Engineering exists at the intersection of cyberworthiness, mission-based cyber risk assessments, graph-based cyber risk assessments and model-based cyber risk assessments. To facilitate the development of a toolkit to implement this hypothesis, we derived some cornerstone principles to guide the toolkit development.

The starting point from these principles was cyberworthiness, as demonstrating cyberworthiness was the goal of any implementation, and the other three components simply provided established techniques to contribute to that

demonstration of cyberworthiness. The natural next step was to leverage the principles for the governance of seaworthiness discussed in Section 2.2:

1. Mindful – the operating context of the system is thoroughly understood during the assessment process
2. Collaborative – all relevant stakeholders engage in joint problem solving during the assessment process
3. Transparent – everyone involved in the assessment process understands what was decided and why
4. Accountable – decision makers are sufficiently informed for them to be accountable for the decisions they make in the assessment process

The first principle of mindfulness was incorporated into Threat-based Cybersecurity Engineering by requiring the explicit modelling of adversary behaviour in the context of the systems operation as the foundation of the approach. This draws on the literature on Mission-based Cyber Risk Assessments, which includes several techniques for incorporating red teams, adversary emulation specialists and other methods for understand system threats into the assessment of the system for a specific mission or set of missions.

The second principle of collaboration was considered, and it was concluded that collaboration in the context of Threat-based Cyber Risk Assessment is about more than simply asking stakeholders for a peer review or involving them in workshops. Collaboration is not possible without an ability to effectively communicate the detail of the assessment amongst the stakeholders, and to do

this the assessment must be documented intuitively, to reduce the barrier of entry for those trying to understand and participate in the assessment. Achieving this within Threat-based Cybersecurity Engineering leverages the literature on Graph-based Cyber Risk Assessments, which provides numerous options for presenting detailed cyber analysis using explainable graphs that can be interpreted intuitively.

The third principle of transparency in the assessment requires that the critical factors of the assessment – threats, assets, controls, risks and decisions – are explicitly traceable to each other and to the assumptions and input variables that drive the analysis of those factors. The detail behind those assumptions and input variables must also be openly documented and made available for critical scrutiny by stakeholders – there must be no opaqueness as to what is driving the results of the assessment. To achieve this transparency, the Model-based Cyber Risk Assessment literature was relied upon, as the model-based approaches provide a structured language built on objects and relationships that natively handles this traceability. They also provide the ability to create a hierarchical assessment, where very specific detail can be defined and simulated to highlight the assumptions and judgements that drive the overall results.

The final principle of accountability is interpreted as informed decision-making, as defined in Section 1.3.5. This principle, in the context of a Threat-based Cybersecurity Engineering approach, is served by the combination of the three other principles. When assessments are conducted in a way that is mindful of the

operating context and an understanding of the threat environment, are detailed with sufficient transparency and specificity to allow stakeholders to understand how the assessment results were determined and are presented in a way that is intuitive and understandable to those without deep cybersecurity expertise, decision makers are provided with the necessary inputs to facilitate informed decision-making.

In addition to these seaworthiness principles, there is also a need for the assessment to stay relevant throughout a system's lifecycle and for effort invested into one assessment to be reused on future assessments to achieve the necessary efficiency to support uptake in professional settings. There are certainly advantages in this respect if the assessment approach is aligned with the contemporary approaches for designing systems, as changes in the design over time can trigger a change to the cyber risk assessment if they are co-located in the same toolset. For this reason, the use of the SysML language from the model-based system engineering discipline was chosen as a cornerstone principle of Threat-based Cybersecurity Engineering.

Ultimately, the following cornerstone principles for Threat-based Cyber Engineering were derived as part of hypothesis development:

- Explicitly model adversary behaviour as a foundation
- Utilise graph-based methods for performing the threat analysis
- Transparently highlight the factors that drive assessment results
- Leverage existing functionality of SysML

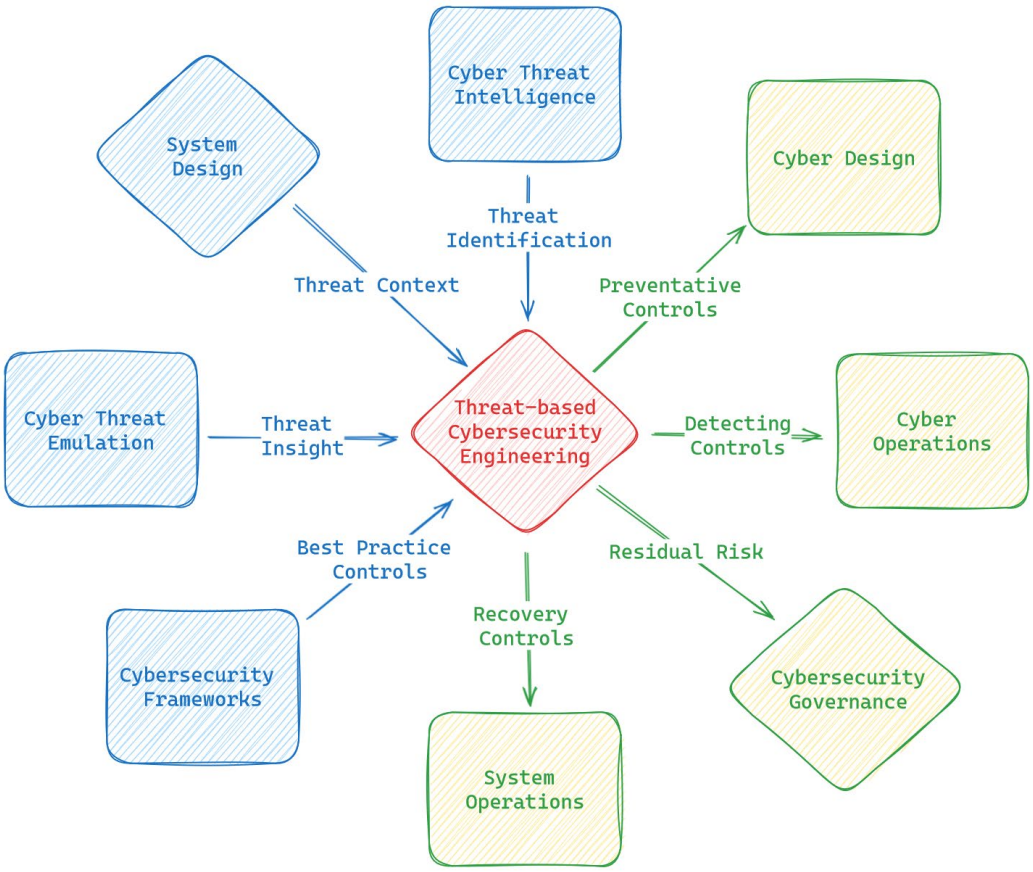


Figure 3-2: *Threat-based Cybersecurity Engineering – Inputs and Outputs*

Now that the guiding principles for the development of a Threat-based Cybersecurity Engineering implementation are understood, we require an enumeration of the inputs and outputs to define the process boundaries to clearly scope the activities that will be performed. The guiding principles provide the 'how' we do Threat-based Cybersecurity Engineering, but now we must discuss 'what' we do as part of Threat-based Cybersecurity Engineering.

Figure 3-2 shows the inputs and outputs to Threat-based Cybersecurity Engineering, allowing us to define the boundaries around the process in terms of the interfaces between other cybersecurity-related functions with an organisation. The blue boxes are those functions that provide inputs to the process, while the

green boxes are those functions that receive outputs from the process. The high-level responsibility in the Threat-based Cybersecurity Engineering function is to provide coordination between those interfacing functions and transform the inputs into the outputs.

The primary inputs to Threat-based Cybersecurity Engineering are:

- Threat Context – taken from the system design
- Threat Identification – leveraging the Cyber Threat Intelligence team
- Threat Insight – utilising Cyber Threat Emulation experts
- Best Practice Controls – derived from the various Cybersecurity Frameworks that are available

The primary outputs from Threat-based Cybersecurity Engineering are:

- Preventative Controls – a baseline of preventative cybersecurity controls for inclusion in the cyber design of a system
- Detecting Controls – a baseline of detection and response controls for implementation by the Cyber Operations team
- Recovery Controls – a baseline of recovery and resilience controls for implementation by the System Operations team
- Residual Risk – the overall risk presented by the threats to the capability given the mitigation mechanisms that are in place

Critically, these inputs and outputs define the scope of Threat-based Cybersecurity Engineering by outlining what it does not do. It does not document

or produce any of the inputs, as it expects those things to have been generated externally, and it does not implement or treat any of the outputs, as it expects those things to be handled externally.

Comparing this Threat-based Cybersecurity Engineering scope to the processes for the contemporary approaches outlined in Section 1.2.1, the scope aligns with Step 2 to Step 5 of the RMF, and Step 1 to Step 4 of the ISM risk framework. However, a governance decision to take a continual approach that applies Threat-based Cybersecurity Engineering throughout the sustainment phases of system lifecycle would address the ongoing monitoring obligations of Step 7 in the RMF and Step 6 in the ISM.

3.2 RESEARCH QUESTIONS

Research questions provide a structure to guide the formal investigation into the research focus outlined in Section 3.1. Research questions have been developed to specify what will be investigated and how success will be measured.

The primary research question for this investigation is:

> *Are integrated threat models, developed using Model-based Systems Engineering tools, an effective and efficient basis for the assessment and evaluation of cyberworthiness in cyber-physical systems?*

This question provides a direct evaluation of the hypothesis outlined in Section 3.1.3. The primary outcome of the investigation is to understand whether the

hypothesis that a threat-focused assessment methodology that leverages model-based system engineering tools demonstrate cyberworthiness, is valid. The research question explicitly identifies the measures with which the assessment process will be judged – namely the efficiency and effectiveness of the process.

There are three secondary research sub-questions beneath the primary research question that provide more specific areas for investigation. These sub-questions primarily enumerate the ways in which the effectiveness and efficiency mentioned in the primary research question will be determined. The first research sub-question is:

> *Do the developed threat models provide decision makers with the necessary understanding to make informed security risk decisions?*

This question highlights that the critical component of effectiveness in the cyberworthiness paradigm is the level of understanding that decision makers have in the assessment, and how that contributes to an informed decision-making process. This reflects the understanding that in a domain that is adversarial, where the likelihoods and input variables can be both unknown and unknowable, the absolute accuracy of the assessment may not be the best measure of effectiveness. Rather, when decision makers are asked to balance a residual risk of a cyber threat determined by a cybersecurity subject matter expert against their understanding of the operational and organisational needs to operate the system being evaluated, the most effective cyber assessment will be the one

where the residual risk is articulated to the decision maker in a way they can understand and appreciate the risks involved.

The result of an effective cyberworthiness risk assessment is an informed decision-making process. Accordingly, this research question is investigating whether the hypothesised solution can meaningfully inform decision makers.

The second research sub-question is:

> *Does the process provide sufficient reusability and maintainability that the methodology is more efficient than prevailing control-focused approaches?*

This question focuses on the efficiency of the hypothesised solution and provides further detail of how the efficiency is perceived and how an acceptable level of efficiency is achieved. The question clarifies that the efficiency of the hypothesised approach should be determined via comparison to the contemporary control-focused approaches outlined in Section 1.2.1. Additionally, the references to reusability and maintainability indicate that the timeframe for this efficiency determination should be more holistic than a single assessment activity. Efficiency of the overall lifecycle of the assessment process is important, including the maintenance and sustainment of a single assessment over time, and the level of reuse between current assessments and future assessments.

The final research sub-question is:

Do cyber security risk practitioners prefer the integrated threat model approach to traditional security risk assessment processes for assessing cyberworthiness?

This question focuses on a user satisfaction component to reinforce the findings of the specific sub-questions focused on effectiveness and efficiency. This question refocuses the primary research question towards practical and actionable outcomes with a specific acknowledgement that the preferences of practitioners, with their built-in biases, are actually an important factor in determining whether a new approach will be implementable in practice. Once again, the user satisfaction of the hypothesised approach will be determined via comparison to the contemporary control-focused approaches outlined in Section 1.2.1.

These research questions provide the guiding principles for the research investigation. The confidence with which the research questions can be answered following the investigation are included in Chapter 7.

3.3 APPROACH

Given the research gaps identified in Section 3.1.1, and the desire to ground the research in practical applications, there was a need for a significant prototype development activity during the early phases of the research program. The decision was also taken to use qualitative research methods, specifically survey questionnaires provided to trial participants, to gather the necessary data to

answer the research questions. This was because the research questions were seeking comparative conclusions, as terms like efficient, effective and informed must be contextualised within current experiences of equivalent cyber security risk assessment methodologies used in practice. While a comparative trial between a contemporary approach and our proposed approach could produce relevant data, it would require additional trials, and it would be difficult to control independent variables between the trials. Accordingly, research methods that surveyed industry experts to extract their qualitative responses to specific questions about the trial they participated in were determined to be the most appropriate approach.

Consequently, the research approach consistent of four phases:

1. Develop a toolkit and implementation methodology that fulfills the hypothesis outlined in Section 3.1.3.
2. Create a survey questionnaire to gather data to help answer the research questions articulated in Section 3.2.
3. Present the toolkit and the theory behind it to industry experts and seek their initial feedback using the survey questionnaire.
4. Apply the toolkit to a real-world system-of-systems in an implementation trial and seek feedback from trial participants using the same survey questionnaire.

In the following sections, each of these four steps will be described in detail.

3.3.1 Prototyping and Development

The first step involved the prototyping and development of a model-based approach to cyberworthiness assessments. The result of this effort was the Cyber Evaluation and Management Toolkit (CEMT), which is a plugin to the commercially available model-based systems engineering software called CATIA Magic [115] that builds upon the diagrams and metamodel provided in the Object Management Group's (OMG) SysML specification [116].

The CEMT was iteratively developed, and early versions were presented at academic conferences [117] and industry working groups to solicit feedback and suggestions from knowledgeable practitioners, from both Australia and internationally, on how to improve the toolkit. Early prototype implementations of the toolkit on small scale industry projects helped to refine the usability of the tool and guide ongoing feature enhancements.

The CEMT development was guided by seven key principles:

1. Threat-focused and threat-first – The CEMT models the behaviour of the adversary as they try to attack the system, rather than describing the system being evaluated, and everything else is derived from that adversary model
2. Reusable – Modelling activities should not be repeated unnecessarily; create the data once and use object referencing and inheritance to reuse as required

3. Intuitive – Relevant assessment information should be presented graphically in simple flow charts that do not require an understanding of SysML to interpret
4. Transparent – Assessment must be documented and easily reviewable, with all assumptions and input variables clearly identified and justified
5. Simulation – Risk assessments should be quantified, with uncertainty, to support detailed simulation and calculation of risk probabilities.
6. Automation – Repetitive and intensive tasks, including both creation and review tasks, should be automated wherever possible to streamline the modelling process
7. Maintainable – Models produced by the CEMT should not be a burden to maintain throughout the system lifecycle and the CEMT process must support the ongoing sustainment of the produced models

This prototyping and development phase ran from 2019 to 2022. The developed CEMT is open-source and freely available online [118]. Further detail of the CEMT is included in Chapter 4.

3.3.2 Survey Questionnaire

The survey questionnaire was developed to gather data from trial participants to help answer the research questions. The questions in the survey can be categorised into six themes:

1. Demographics – to determine respondent's experience and relationship to the assessment process
2. Participation Validation – to ensure the respondents understood the assessment process they were being asked to rate
3. Efficiency – to identify whether the assessment process was considered efficient by respondents
4. Effectiveness – to identify whether the assessment process was considered effective by respondents
5. Decision Making – to identify whether respondents thought the decisions made as part of the assessment process were well-informed
6. General Opinions – to gather general feedback on the assessment process

The demographic themed questions from the survey questionnaire are listed in Table 3-1. These questions are intended to capture background information about the respondent to help categorise the dataset based on the role and experience level of the respondents.

Table 3-1: *Demographic questions in the survey questionnaire*

Demographics	
Question	**Response Format**
Which option best describes your organisation?	Enumeration[1]
What best describes your relationship with the development of cyber/physical system(s)?	Enumeration[2]

[1] Enumeration Options – Industry | Government | Defence
[2] Enumeration Options – Technical Specialist | Management | Governance

Demographics	
Question	**Response Format**
What is your main area of expertise?	Enumeration[3]
Do you have a background as a cybersecurity subject matter expert?	Yes / No
How many years of experience do you have as a cybersecurity subject matter expert?	Integer
Do you have experience working with cybersecurity risk assessments, either as a subject matter expert, reviewer or stakeholder?	Yes / No
How many years of experience do you have working with cybersecurity risk assessments?	Integer
What option best describes the role you performed in the cyberworthiness assessment process?	Enumeration[4]
Have you participated in a cyberworthiness assessment process previously?	Yes / No

The questions designed to validate the respondent's participation in the assessment process are shown in Table 3-2. These questions were included to confirm whether the respondents understood the assessment process, their role in that assessment process and the Cyber Evaluation and Management Toolkit (CEMT) used to conduct the assessment. The 5-point Likert scale used in this series of questions was Strongly Agree, Agree, Neutral, Disagree and Strongly Disagree.

[3] Enumeration Options – IT/Computing | Engineering | Management | Operations | Maintenance
[4] Enumeration Options – Assessor/Modeller | Peer Reviewer (Cyber Expertise) | Peer Reviewer (Non-Cyber Expertise) | Decision Maker | System Designer/Technical Lead | End User

Table 3-2: *Participation validation questions in the survey questionnaire*

Participation Validation	
Question	**Response Format**
Sufficient explanation of the CEMT was provided	5-point Likert
The role I had in the cyberworthiness assessment process using the CEMT was clearly understood	5-point Likert

The questions related to the effectiveness of the assessment process are listed in Table 3-3. These questions cover the accuracy of the assessment, the transparency of the assessment and the ability of the assessment process to provide critical insights in the cyberworthiness of the system. The intent of these questions is to gather data from the respondents that can be analysed to identify the comparative effectiveness of the assessment process against contemporary approaches. The 5-point Likert scale used in this series of questions was Strongly Agree, Agree, Neutral, Disagree and Strongly Disagree.

Table 3-3: *Effectiveness-related questions in the survey questionnaire*

Effectiveness	
Question	**Response Format**
The cyberworthiness assessments produced by the CEMT are sufficiently detailed	5-point Likert
The CEMT clearly identifies the factors which contribute to the cyberworthiness of a system	5-point Likert
The CEMT clearly identifies which security controls are important to the system	5-point Likert
The CEMT clearly identifies why particular security controls are important	5-point Likert
The CEMT provides a complete and comprehensive approach to determining cyberworthiness	5-point Likert

Effectiveness	
Question	**Response Format**
The CEMT produces accurate assessments of a system's cyberworthiness	5-point Likert
The CEMT identifies additional security controls which are likely to provide the biggest improvement to the cyberworthiness of a system	5-point Likert
The CEMT identifies the relative impact of security controls with respect to the cyberworthiness of the system	5-point Likert
The CEMT produces transparent cyberworthiness assessments	5-point Likert
The CEMT produces risk assessments that are tailored to the context in which the system operates	5-point Likert
The CEMT produces cyberworthiness assessments that have ongoing value through the future phases of the capability life cycle	5-point Likert
The CEMT clearly highlights the areas of greatest cyber risk to the system	5-point Likert

The questions related to the efficiency of the assessment process are listed in Table 3-4. These questions cover the time taken to complete the assessment process, the complexity of the assessment process and friction around stakeholder engagement and peer review. The intent of these questions was to gather data from the respondents that can be analysed to identify the comparative efficiency of the assessment process against contemporary approaches. The 5-point Likert scale used in this series of questions was Strongly Agree, Agree, Neutral, Disagree and Strongly Disagree.

Table 3-4: *Efficiency-related questions in the survey questionnaire*

Efficiency	
Question	**Response Format**
The CEMT clearly scopes the context in which risk assessments are made.	5-point Likert
Cyberworthiness assessments are simple to produce using the CEMT	5-point Likert
The CEMT provides a good balance between the quality of the cyberworthiness assessment and the time to produce that assessment	5-point Likert
The CEMT is an effective use of time	5-point Likert
The CEMT process is intuitive	5-point Likert
The CEMT is not overly dependent on the subjective opinion of subject matter experts	5-point Likert
The CEMT allows subject matter experts to justify their cyberworthiness recommendations	5-point Likert
The CEMT facilitates the engagement of stakeholders and the provision of meaningful input from those stakeholders into a cyberworthiness assessment	5-point Likert
The CEMT encourages stakeholders to work collaboratively to determine the residual risk level	5-point Likert
The CEMT presents cyberworthiness assessments in a way that promotes collaboration	5-point Likert
The CEMT assessments are easily adaptable to changes in the system design	5-point Likert
The CEMT assessments can be easily modified to capture additional threat vectors identified during review workshops	5-point Likert
The CEMT assessments can be easily kept up to date as the system design matures throughout the system lifecycle	5-point Likert
The CEMT assessments are agile and responsive to changes	5-point Likert
A CEMT-based approach to cyberworthiness evaluations supports evergreen development of complex systems	5-point Likert

The questions related to the ability of the assessment process to support informed decision-making are listed in Table 3-5. These questions cover the assessment process's ability to convey the risk to a decision maker in a way they can understand, its ability to provide insight into the cyberworthiness of the system and the sufficiency of the information provided to the decision maker. The intent of these questions was to gather data from the respondents that can be analysed to identify whether the assessment process supports informed decision-making. The 5-point Likert scale used in this series of questions was Strongly Agree, Agree, Neutral, Disagree and Strongly Disagree.

Table 3-5: *Decision-related questions in the survey questionnaire*

Decision Making	
Question	**Response Format**
The CEMT makes it clear why risk decisions were made	5-point Likert
The CEMT makes it clear how risk decisions were made	5-point Likert
The CEMT facilitates informed decision making with respect to the identified cybersecurity risks	5-point Likert
Cyberworthiness assessments produced with the CEMT are easy to understand	5-point Likert
The CEMT presents cyberworthiness risk to the system in an understandable manner	5-point Likert
The CEMT provides new insights in the cyberworthiness of a system	5-point Likert
The CEMT facilitates effective decisions and/or recommendations regarding the cyberworthiness of the system	5-point Likert
The CEMT provides sufficient information to allow decision makers to be accountable for their decisions	5-point Likert
The CEMT presents information to decision makers in a manner that allows them to make the right decision	5-point Likert

In addition to the specific questions about the effectiveness and efficiency of the assessment process, the survey question also included general questions that allowed the respondents to convey their general feelings about the assessment process. These questions are documented in Table 3-6. The intent of these questions was to gather data from the respondents about their subjective impression of the overall assessment process. The 5-point Likert scale used in this series of questions was Strongly Agree, Agree, Neutral, Disagree and Strongly Disagree; except for the notated question, which used a different scale.

Table 3-6: *General questions in the survey questionnaire*

General Opinions	
Question	**Response Format**
Overall, what rating would you give the CEMT as a cyberworthiness assessment methodology?	5-point Likert[5]
The CEMT adds value to a system and/or project	5-point Likert
The CEMT is an improvement over existing cyberworthiness assessment processes	5-point Likert
A comprehensive cyberworthiness assessment is a critical component of the system capability lifecycle	5-point Likert
The CEMT improves my understanding of the cyberworthiness of a system	5-point Likert
The CEMT is an appropriate technique for assessing cyberworthiness on systems you work with	5-point Likert
How likely are you to recommend the application of CEMT in one of your future projects/systems	Enumeration[6]

[5] Likert Scale – Poor | Fair | Good | Very Good | Excellent
[6] Enumeration Options – Any Integer between 0 and 10

The questions in the survey questionnaire were reviewed by the Defence People Research – Low Risk Ethics Panel (DPR-LREP) and ethics approval was provided to conduct the research using these questions via online survey or focus group panels. This approval was provided under protocol *061-21: Cyberworthiness of Cyber-Physical Systems*, which was subsequently ratified by the Human Research and Ethics Committee (HREC) at the University of New South Wales.

3.3.3 Face Validity Trial

Face validity trials were conducted on the CEMT to obtain confidence that implementation trials of the toolkit would be able to sufficiently answer the research questions. Face validity is used in this context to mean a hypothesised validity based on the intuition of a group of domain experts. Face validity has been used in psychological research dating back to 1947 [119], but it should be noted that its use is also considered controversial [120].

This controversy surrounding the legitimacy of face validity measures will not be explored in further detail, because the face validity trials in this methodology are not being used to judge the overall validity of the hypothesis, but rather is used to gain confidence that the prototyping and development phase has produced a process that is sufficiently mature to proceed to implementation trials in order to formally determine validity. Implementation trials on complex cyber-physical systems can be costly, both in terms of time and money. A face validity trial simply

allowed us to complete an independent, surface-level evaluation of the maturity of the toolkit prior to committing to a full implementation trial.

The face validity trial was conducted between 2022 and 2023 and consisted of the researchers presenting the CEMT and associated processes at academic conferences, tutorials and industry working groups within Australia and internationally. Attendees at these presentations were invited to anonymously complete the survey questionnaire described in Section 3.3.2, with the intent to understand whether those attendees hypothesised that the CEMT would be able to address the themes covered by the survey questionnaire.

The face validity trial gave us confidence to proceed to implementation trials and helped to convince system owners to allow us to apply the CEMT to their complex cyber-physical systems. The results of the face validity trial are discussed in detail in Chapter 5.

3.3.4 Implementation Trial

Implementation trials using the CEMT were undertaken by applying the assessment process to a complex cyber-physical system to gather data to address the research questions and validate the hypothesis. The implementation trial of the CEMT was conducted on a complex Australian naval capability [26] between March 2024 and November 2024. This trial covered the entire naval capability, which included IT, OT and integrated cyber-physical systems.

The implementation trial involved an application of the CEMT by a collaborative group of cybersecurity experts, system experts and researchers to conduct a comprehensive assessment of the naval capability. Following the completion of the assessment, the modellers, reviewers, designers and decision makers were all invited to anonymously complete the survey questionnaire described in Section 3.3.2.

The implementation trial helped to validate the hypothesis and answer the research questions. The results of the implementation trial are discussed in detail in Chapter 6.

3.4 HYPTHESISED CONTRIBUTION

Completing this research is expected to contribute the following knowledge to the literature:

1. A novel methodology for completing threat-focused cyber risk assessments of cyber-physical systems that leverages model-based systems engineering techniques will be designed and created
2. An open-source, freely available implementation of the methodology will be developed and published
3. Insight, based on practical trials, into how well this threat-focused methodology can evaluate the cyberworthiness of cyber-physical systems in terms of both effectiveness and efficiency

4. Practitioner-derived learnings about the ability for these types of assessments to lead to more informed cyber risk decision making in real-world situations
5. An understanding, based on practical trials, of the strengths and weaknesses of threat-focused cyber risk assessments in comparison to control-focused cyber risk assessments on complex cyber-physical systems

These contributions would make meaningful progress towards closing the literature gaps identified in Chapter 2.

Beyond that, this research is also expected to make a significant contribution to the professional domain of cybersecurity engineering, as a demonstrated practical application against a large, complex system can provide the confidence and justification for the uptake of the same or similar threat-focused methodologies that leverage model-based systems engineering techniques across the industry. As identified in Section 1.2.3 there is a strong need for more efficient and effective methodologies for the assessment of cybersecurity risk with complex technology systems, particularly within the Defence industry. A practical and pragmatic approach that has been demonstrated as feasible within this industry has the potential to become a new normal within that industry, which could lead to significant cost savings to taxpayers through improved efficiency and concrete national security benefits through improved effectiveness.

CHAPTER 4 CEMT OVERVIEW

4.1 BACKGROUND

This chapter provides a detailed explanation of the Cyber Evaluation and Management Toolkit (CEMT) which was created during the Prototyping and Development phase of the research methodology outlined in Chapter 3 to address the literature gaps articulated in Chapter 2. The CEMT is a plugin to the commercially available model-based system engineering tool called *CATIA Magic Cyber Systems Engineer*. The CEMT plugin consists of a profile that extends the SysML stereotypes that are included in the *CATIA Magic Cyber Systems Engineer* software package. This chapter details the taxonomy of the

CEMT and provides examples of the how they are used to develop, analyse and assess the cyber threats to a cyber-physical system or capability. The examples within this chapter model simple threats to help explain the overall process of producing a model. Additional examples of attacks against cyber-physical systems can be found in the provided sample model [118].

4.1.1 CEMT Process

The CEMT process can be broken down into three cardinal phases, the *Threat Modelling* phase, the *Threat Allocation* phase and the *Risk Assessment* phase. These phases, shown in Figure 4-1, will be used as the structure of this chapter to categorise and explain the taxonomy of the CEMT.

Figure 4-1: *The three cardinal phases of the CEMT process*

The three cardinal phases consist of the following activities:

1. *Threat Modelling* – A detailed model of potential adversary actions is developed along with potential mitigations against those actions.
2. *Threat Allocation* – The security-relevant assets within the system are identified and a threat-relevant control baseline is derived from the threat model.

3. *Risk Assessment* – Critical threat vectors from the threat model are quantitatively simulated to calculate the likelihood of the risk associated with that threat vector, which is combined with the consequence of the attack to determine a risk rating.

While the phasing shown in Figure 4-1 provides a high-level overview of the phases of the CEMT process, a detailed application of the CEMT on a real-world system will be significantly more iterative than a three-step sequential process. In Section 4.5, a more comprehensive CEMT process flow will be outlined that takes these three cardinal phases and incorporates them into a more holistic, iterative and implementable process for the application of the CEMT on complex cyber-physical systems.

4.1.2 Purpose and Focus

The CEMT is primarily focused on providing assurance of the likelihood of a given threat or risk and articulating that in a manner that can be consistently applied by security assessors, meaningfully reviewed by subject matter experts and clearly understood by decision makers. The CEMT does not aid in the initial identification of risks or threats to a cyber-physical system, and it does not assist in determining the consequence of that risk or threat being realised.

The top-level threats to the system need to be determined and captured in the model, but the CEMT does not include or provide any tools, techniques or guidance to assist in the identification of these threats. Existing threat

identification methods must be used to generate a list of top-level threats to the system that is being evaluated with the CEMT. Similarly, the consequence of a particular risk or threat is captured in the tool, but the focus of the CEMT is in determining the likelihood of a risk, not the consequence. An external criticality assessment should be conducted on each asset to inform the consequence of a risk or threat to that asset in the context of the broader functional or organisational capability.

This focus on providing assurance of the likelihood value is a deliberate choice. Discussions regarding criticality and consequence of cybersecurity risks are typically less opaque and less subjective, while top-level threats to most systems are typically well documented in best-practice literature and therefore reasonably well understood. Traditionally, it is the likelihood determination where there is most subjectivity and opaque reasoning, and hence this determination is the primary focus of the CEMT.

4.2 THREAT MODELLING

The *Threat Modelling* phase requires the development of a detailed model of potential malicious behaviour against the system being evaluated. These behaviours are categorised into Misuse Cases and described using nested Mal-Activities that provide a flow chart of adversary actions against the system and the ability of the system to detect those actions.

4.2.1 Misuse Cases

Misuse Cases are analogous to the system's functional use cases, which describe the functions of a system and the actors that participate in delivering those functions. Misuse cases, however, represent the high-level attacks being performed by an adversary against a system to deliver the adversary's desired outcomes – which are typically to compromise the system's confidentiality, integrity or availability to produce a tactical or strategic effect.

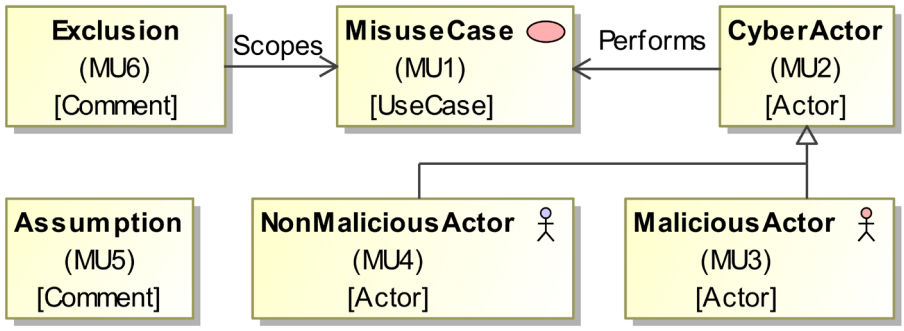


Figure 4-2: *CEMT Misuse Case Taxonomy*

Figure 4-2 depicts the taxonomy of stereotypes that the CEMT includes for the development of Misuse Cases:

MU1. `MisuseCase` – This stereotype inherits from the built-in `UseCase` metaclass and is used to represent the misuse cases defined by the modeller.

MU2. `CyberActor` – This stereotype inherits from the built-in Actor metaclass and is a generic stereotype used to group the various actors that participate in the defined misuse cases. Objects stereotyped by `CyberActor` can perform a `MisuseCase`.

MU3. `MaliciousActor` – This stereotype inherits from the `CyberActor` stereotype and represents any malicious actor that the modeller defines as performing a `MisuseCase`, such as an external adversary or malicious insider.

MU4. `NonMaliciousActor` – This stereotype inherits from the `CyberActor` stereotype and represents any non-malicious actor that the modeller defines as performing a `MisuseCase`, such as a system operator or maintainer.

MU5. `Assumption` – This stereotype inherits from the built-in `Comment` metaclass and is used by the modeller to articulate any assumptions made during the CEMT assessment. It is primarily used in the *Threat Modelling* phase but can be used on any CEMT diagram.

MU6. `Exclusion` – This stereotype inherits from the built-in `Comment` metaclass and is used by the modeller to scope a `MisuseCase`, such as clarifying whether a military overrun would be considered a Physical Incursion.

These stereotypes are used by the modeller to produce *CEMT Misuse Case Diagrams*, which are custom diagrams within the CEMT that are based on built-in *SysML Use Case Diagrams*. An example of a *CEMT Misuse Case Diagram* is shown in Figure 4-3.

The *CEMT Misuse Case Diagram* is centred around the `MisuseCase` objects, which represent the high-level threats to the system being evaluated. In Figure

4-3, *External Network Penetration*, *Insider Threat* and *Physical Incursion* are all examples of `MisuseCase` objects. The *CEMT Misuse Case Diagram* also includes `MaliciousActor` objects (*External Adversary* and *Malicious Insider*) and `NonMaliciousActor` objects (*Operator* and *Maintainer*) that are linked to the misuse cases they participate in via a built-in `Association` relationship.

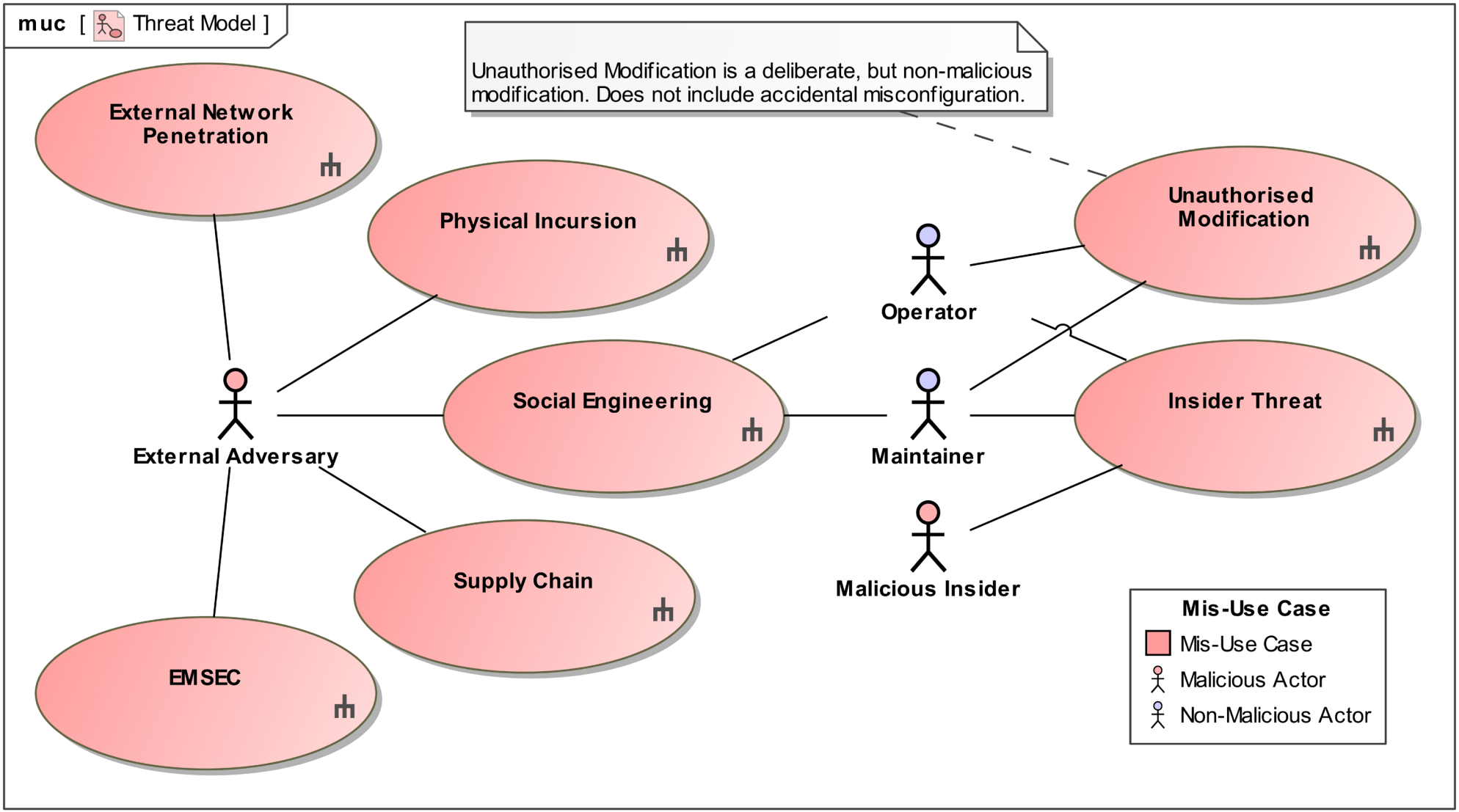


Figure 4-3: *Sample CEMT Misuse Case Diagram*

4.2.2 Mal-Activities

While the misuse cases described in Section 4.2.1 perform a scoping exercise for the development of the threat model by enumerating the breadth of potential adversarial cyber-attacks, the mal-activities covered in this section provide the specific detailed behaviour that underlies each misuse case.

Mal-activities are analogous to system functional behaviour models, which use activity diagrams to produce a nested flow chart of how actions performed by different components within a system-of-systems can combine to produce emergent functional behaviour by the overall capability. Mal-activities, however, describe the behaviour of malicious actors using a series of nested flow charts, to enumerate the various ways in which that malicious actor can compromise the confidentiality, integrity or availability of the system or capability being evaluated.

Mal-activities are subordinate to the misuse cases that they describe, and they contain actions which describe the individual steps that detail the behaviour represented by the mal-activity. The actions are connected by flows which define the sequence and order by which the actions must be performed by the adversary to complete the behaviour of the mal-activity.

The mal-activities in the CEMT also describe the actions that a system and its operators can take to detect the malicious actor as they perform their attack. This allows security mitigations that focus on detecting attackers to be explicitly modelled as part of the *Threat Modelling* phase.

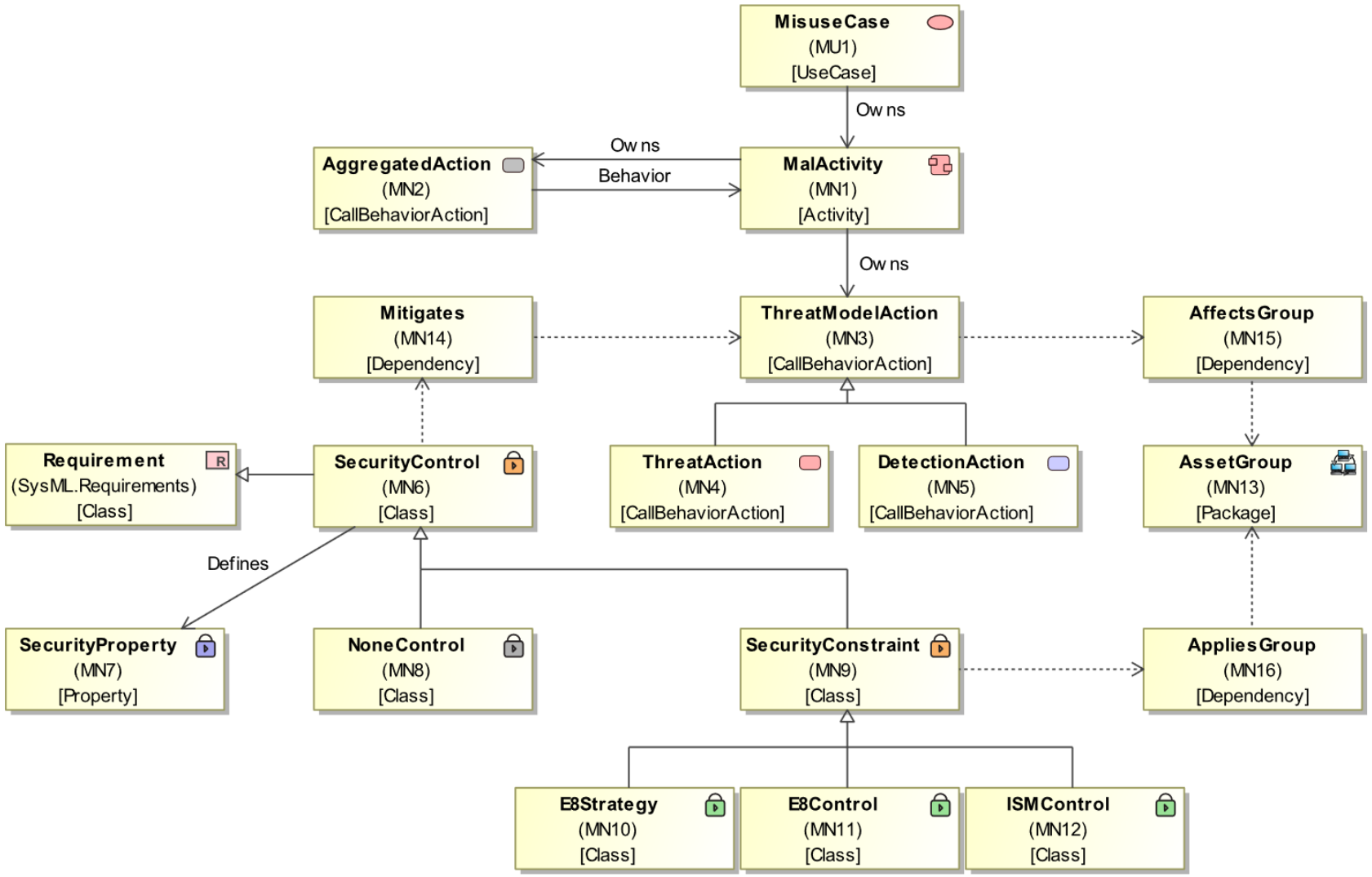


Figure 4-4: *CEMT Mal-Activity Node Taxonomy*

Figure 4-4 shows the taxonomy of stereotypes that the CEMT includes for defining the nodes within Mal-Activities:

MN1. `MalActivity` – This stereotype inherits from the built-in `Activity` metaclass, either classifies a `MisuseCase` (when this is a top-level mal-activity) or defines the behaviour of an `AggregatedAction` (when this is a nested mal-activity) and owns the actions that describe make up the threat model.

MN2. `AggregatedAction` – This stereotype inherits from the built-in `CallBehaviorAction` metaclass and is used to create the intermediate actions that facilitate nested mal-activity diagrams. They are owned by the higher level `MalActivity` and have their behaviour defined by a second lower level `MalActivity`.

MN3. `ThreatModelAction` – This stereotype inherits from the `CallBehaviourAction` metaclass and is used as a parent stereotype for all actions that describe the lowest level of the threat model. This stereotype can be linked to an `AssetGroup` via the `AffectsGroup` dependency and can be linked from a `SecurityControl` via the `Mitigates` dependency.

MN4. `ThreatAction` – This stereotype inherits from the `ThreatModelAction` stereotype and is used to represent the lowest level of adversary action in the threat model.

MN5. `DetectionAction` – This stereotype inherits from the `ThreatModelAction` stereotype and is used to represent the lowest level of detection action in the threat model.

MN6. `SecurityControl` – This stereotype inherits from the SysML `Requirement` stereotype and is used to represent any preventative or detecting control within the threat model. This stereotype can be linked to a `ThreatModelAction` via the `Mitigates` dependency.

MN7. `SecurityProperty` – This stereotype inherits from the built-in `Property` metaclass and is an instantiation of a `SecurityControl`. A `SecurityProperty` carries an implementation state and is specific to one component within the system.

MN8. `NoneControl` – This stereotype inherits from the `SecurityControl` stereotype and is used to create a special `SecurityControl` that represents the lack of any potential security controls. This `NoneControl` is linked to any `ThreatModelAction` that cannot be mitigated by any controls to differentiate an action that has been assessed as having no feasible controls from one which has not yet been assessed.

MN9. `SecurityConstraint` – This stereotype inherits from the `SecurityControl` stereotype and is used to represent security controls which are mandated for assessment via regulation or requirement. A `SecurityConstraint` can be linked directly to an `AssetGroup` via the `AppliesGroup` dependency, which allows a `SecurityConstraint` to

be allocated directly to an `AssetGroup`, rather than it being indirectly related via a `ThreatModelAction` like a standard `SecurityControl`.

MN10. `E8Strategy` – This stereotype inherits from the `SecurityConstraint` stereotype and is used to represent the eight strategies from the Essential Eight Maturity Model [15].

MN11. `E8Control` – This stereotype inherits from the `SecurityConstraint` stereotype and is used to represent the various mitigations that determine the maturity levels of the eight strategies from the Essential Eight Maturity Model [15].

MN12. `ISMControl` – This stereotype inherits from the `SecurityConstraint` stereotype and is used to represent the controls within the Australian Government Information Security Manual [14].

MN13. `AssetGroup` – This stereotype inherits from the built-in `Package` metaclass and is used to represent a specific set of attributes of a given component or `Asset` that would determine whether a `ThreatModelAction` would affect that component or `Asset`.

MN14. `Mitigates` – This stereotype inherits from the built-in `Dependency` metaclass and is used as a relation between a `SecurityControl` and a `ThreatModelAction` to show that the `SecurityControl` is relevant to the `ThreatModelAction` because it could potentially mitigate that step in the threat model if it were to be implemented.

MN15. `AffectsGroup` – This stereotype inherits from the built-in `Dependency` metaclass and is used as a relation between a `ThreatModelAction` and an `AssetGroup` to show that the `ThreatModelAction` is relevant to the `AssetGroup` because this step in the threat model would be performed against an `Asset` or component which has the attributes or properties described by the `AssetGroup`.

MN16. `AppliesGroup` – This stereotype inherits from the built-in `Dependency` metaclass and is used as a relation between a `SecurityConstraint` and an `AssetGroup` to show that the `SecurityConstraint` is relevant to the `AssetGroup` because this mandatory mitigation would be applied to an `Asset` or component which has the attributes or properties described by the `AssetGroup`.

These nodes are used to build the *CEMT Mal-Activity Diagrams* that describe the behaviour of malicious adversaries, identify the security mitigations that are relevant to each step in the attack path and the attributes of an asset of component that would cause them to be the target of each step in the attack path. However, the mal-activity nodes are only half of the equation when it comes to building the *CEMT Mal-Activity Diagrams*, as the nodes need to be connected to show the sequence in which the various nodes need to be performed for an overall risk to be realised and to identify the impact of that risk being realised.

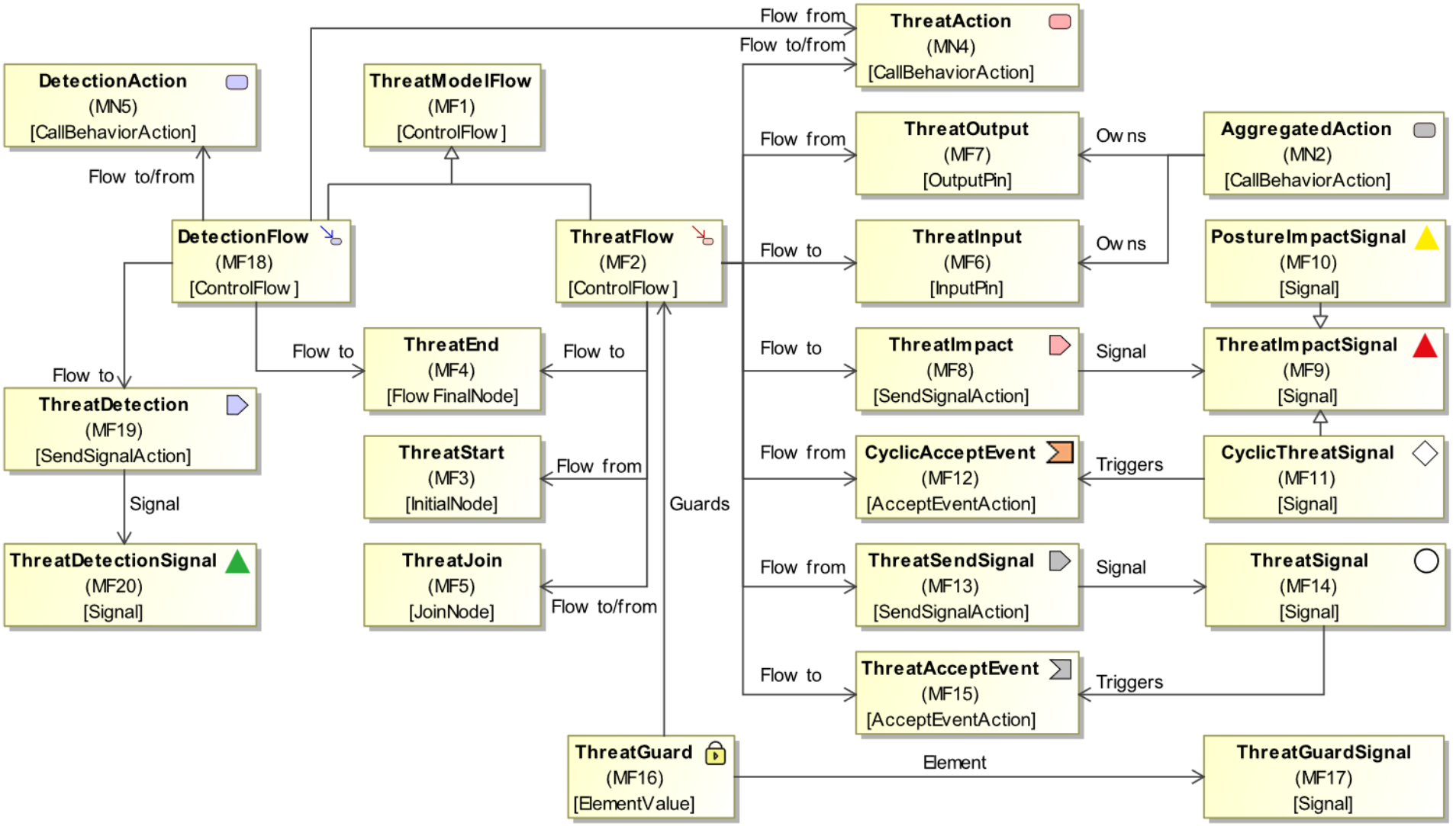


Figure 4-5: *CEMT Mal-Activity Flow and Signal Taxonomy*

Figure 4-5 shows the taxonomy of stereotypes that the CEMT includes for defining the flows and signals between nodes within Mal-Activities:

MF1. `ThreatModelFlow` – This stereotype inherits from the built-in `ControlFlow` metaclass and is used as a parent stereotype for all flows between nodes in the mal-activity.

MF2. `ThreatFlow` – This stereotype inherits from the `ThreatModelFlow` stereotype and represents the path taken by the malicious adversary as they perform their attack. A `ThreatFlow` can flow to and from the different nodes in a *CEMT Mal-Activity Diagram*, as shown in Figure 4-5.

MF3. `ThreatStart` – This stereotype inherits from the built-in `InitialNode` metaclass and represents the starting point of an attack by the malicious adversary. There is typically only one `ThreatStart` per `MisuseCase`, which exists in the top-level mal-activity for that `MisuseCase`.

MF4. `ThreatEnd` – This stereotype inherits from the built-in `FlowFinalNode` metaclass and represents the end of a flow where an adversary has either been prevented from progressing with their attack, or where a detection has been unsuccessful.

MF5. `ThreatJoin` – This stereotype inherits from the built-in `JoinNode` metaclass and is used when two `ThreatFlows` need to be joined together to show that both paths must be activated for the threat to progress along the `ThreatFlow` leaving the `JoinNode`.

MF6. `ThreatInput` – This stereotype inherits from the built-in `InputPin` metaclass and is used as the input pin for an `AggregatedAction`, allowing a `ThreatFlow` to enter the `MalActivity` that defines the `AggregatedAction`. A `ThreatInput` is fundamental to producing a nested set of mal-activities.

MF7. `ThreatOutput` – This stereotype inherits from the built-in `OutputPin` metaclass and is used as the output pin for an `AggregatedAction`, allowing a `ThreatFlow` to leave the `AggregatedAction` and return to the parent `MalActivity`. A `ThreatOutput` is fundamental to producing a nested set of mal-activities.

MF8. `ThreatImpact` – This stereotype inherits from the built-in `SendSignalAction` metaclass and is used to represent the end of a `ThreatFlow` where the overall attack path terminates because the attacker has achieved a reduction in the security posture of the system or successfully completed an attack.

MF9. `ThreatImpactSignal` – This stereotype inherits from the built-in `Signal` metaclass and defines signals that represent a risk being realised at the end of an attack. These are attached as the signal of a `ThreatImpact`.

MF10. `PostureImpactSignal` – This stereotype inherits from the built-in `Signal` metaclass and defines signals that represent a reduction of security posture at the end of an attack. These are attached as the signal of a `ThreatImpact` and are typically used to show the end of one attack

path that requires two or more parallel attacks to be completed simultaneously for the risk to actually be realised – for example, the compromise of a component that has a functional backup in place.

MF11. `CyclicThreatSignal` – This stereotype inherits from the built-in `Signal` metaclass and defines signals that indicate an intermediate signal in threat model that would cause an infinitely cyclic loop if it was not terminated with a `CyclicThreatSignal`. For example, a network penetration attack would usually include the ability to pivot between different devices on the network, at which point the attacker repeats their attack on the next network device, and the `CyclicThreatSignal` allows us to manually terminate and restart the attack path to force the attack tree created by the threat model to be acyclic to avoid infinite loops.

MF12. `CyclicAcceptEvent` – This stereotype inherits from the built-in `AcceptEventAction` metaclass and represents the continuation point for a flow that has reached a `CyclicThreatSignal`. A `CyclicAcceptEvent` is triggered by a `CyclicThreatSignal`.

MF13. `ThreatSendSignal` – This stereotype inherits from the built-in `Signal` metaclass and is used to represent an intermediate node where the overall attack path reaches a point where the flow needs to leave the current *CEMT Mal-Activity Diagram* and continue on another diagram.

MF14. `ThreatSignal` – This stereotype inherits from the built-in `Signal` metaclass and defines signals that indicate an intermediate signal in threat model. These are attached as the signal of a `ThreatSendSignal`.

MF15. `ThreatAcceptEvent` – This stereotype inherits from the built-in `AcceptEventAction` metaclass and represents the continuation point for a flow that has reached a `ThreatSignal`. A `ThreatAcceptEvent` is triggered by a `ThreatSignal`.

MF16. `ThreatGuard` – This stereotype inherits from the built-in `ElementValue` stereotype and is placed as a guard on a `ThreatFlow` to show that the `ThreatGuardSignal` that is the element of the `ThreatGuard` needs to have been triggered for the attack to progress down that flow. This can be used to describe the need for pre-requisites – for example, that the attacker has access to login credentials – to progress down a particular `ThreatFlow`.

MF17. `ThreatGuardSignal` – This stereotype inherits from the built-in `Signal` metaclass and defines pre-requisites or co-requisites that are attached as the element of `ThreatGuard` for use in guard conditions throughout the threat model.

MF18. `DetectionFlow` – This stereotype inherits from the `ThreatModelFlow` stereotype and represents the path that depicts how a system might detect the actions of the malicious adversary. A `DetectionFlow` can flow to and from the different nodes in a *CEMT Mal-Activity Diagram*, as shown in Figure 4-5..

MF19. `ThreatDetection` – This stereotype inherits from the built-in `SendSignalAction` metaclass and is used to represent the end of a

`DetectionFlow` where the adversary attack has been successfully detected.

MF20. `ThreatDetectionSignal` – This stereotype inherits from the built-in `Signal` metaclass and defines signals that represent an attack being successfully detected. These are attached as the signal of a `ThreatDetection`.

These stereotypes are used by the modeller to produce *CEMT Mal-Activity Diagrams*, which are custom diagrams within the CEMT that are based on built-in *SysML Activity Diagrams*. These CEMT Mal-Activity Diagrams are created in a nested hierarchy, from the top-level where each misuse case is described to the bottom-level where the individual steps taken by an attacker are depicted.

Examples of top-level *CEMT Mal-Activity Diagrams* are shown in Figure 4-6 and Figure 4-7. Figure 4-6 describes the *Insider Threat* misuse case and shows a `ThreatStart` for an insider threat connected to a series of `AggregatedActions` which lead to separate `ThreatImpacts` for an *Integrity Loss*, *Availability Loss* and *Confidentiality Loss*. These nodes are linked via `ThreatFlows` to describe the steps a malicious insider would take to compromise the system. Also modelled are the `ThreatAcceptEvent` nodes named *System Access* and *Room Access*, which can be triggered by a `ThreatSendSignal` with a matching signal. Figure 4-7 is the equivalent mal-activity drawing for the *Physical Incursion* misuse case, which includes a `ThreatSendSignal` with a matching *Room Access* and *System Access* signal.

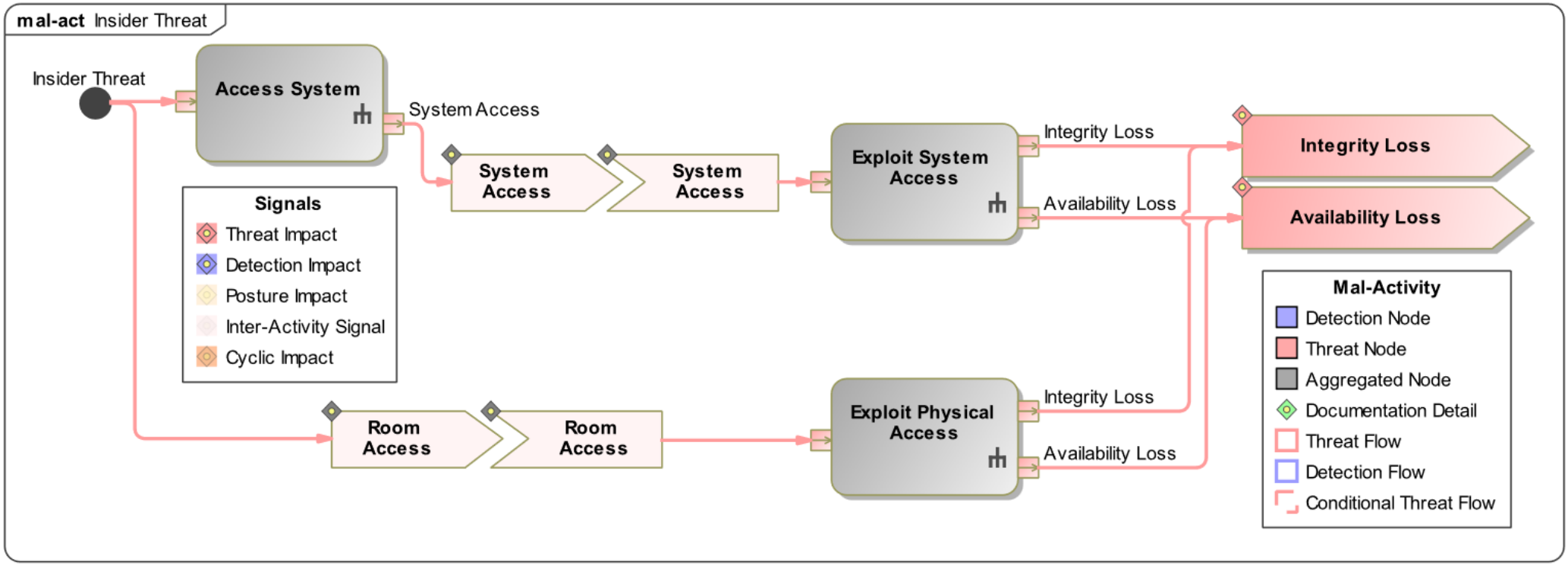


Figure 4-6: *Sample CEMT Mal-Activity Diagram – Insider Threat*

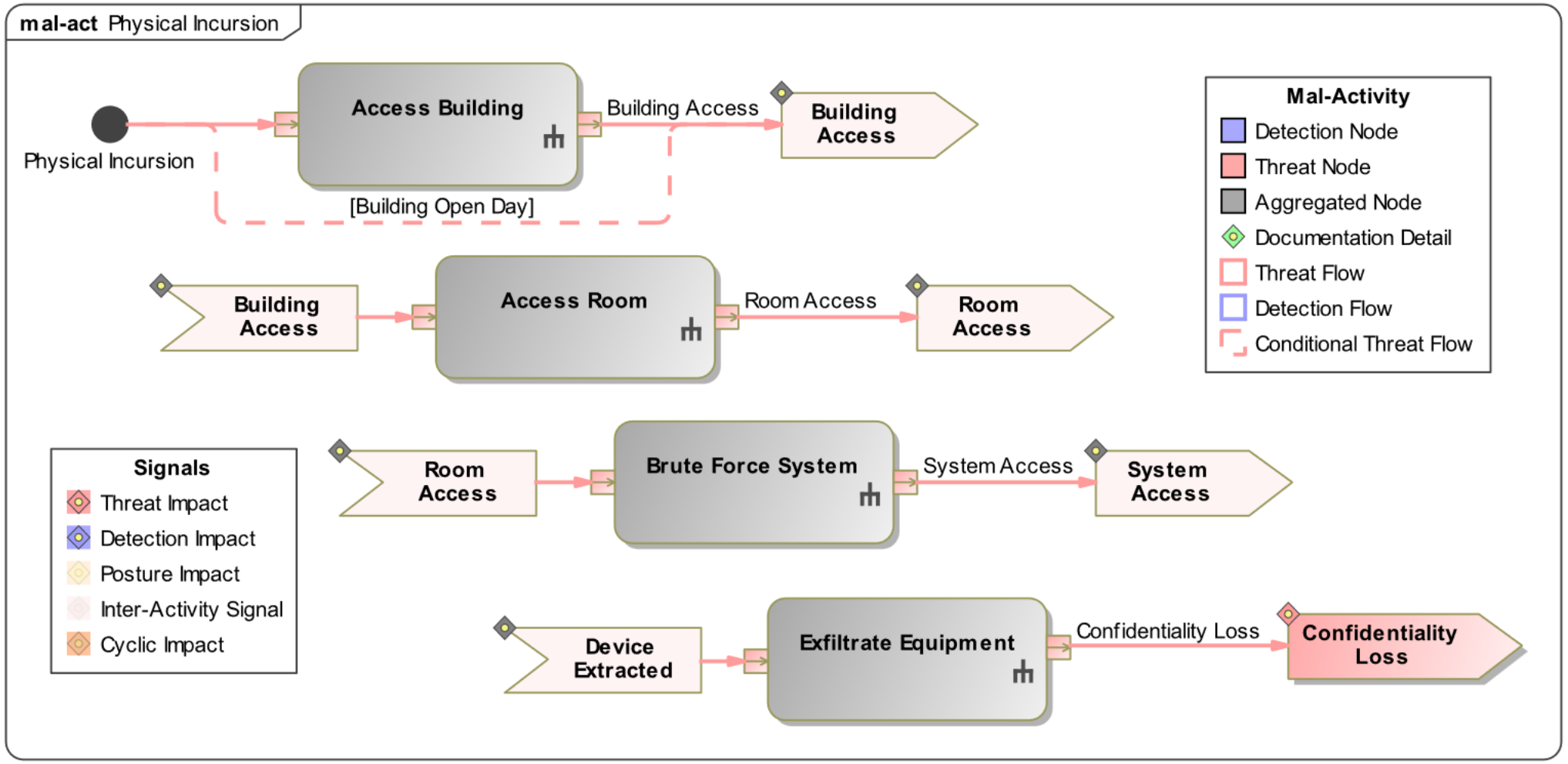


Figure 4-7: *Sample CEMT Mal-Activity Diagram – Physical Incursion*

Figure 4-7 also provides an example of a `ThreatGuard` called *Building Open Day*. This shows that if the pre-requisite condition of the facility having a public open day, where unauthorised personnel are given access to the building, then the Access Building action can be skipped. This allows for the cyber risk during those specific situations to be assessed and analysed alongside the baseline security posture.

Figure 4-8 displays a *CEMT Mal-Activity Diagram* for the *Exploit System Access* `MalActivity`, which describes the detail below the *Exploit System Access* `AggregatedAction` from Figure 4-6. The in and out ports on the border of Figure 4-8 are associated with the `ThreatInput` and `ThreatOutput` pins shown on the *Exploit System Access* `AggregatedAction` in the higher level mal-activity diagram (Figure 4-6).

The intermediate level mal-activity diagram contains an additional layer of detail by enumerating the various ways in which system access could be exploited. These intermediate layers can be repeated to create a deeper hierarchy if it is warranted by the complexity of the misuse case that is being modelled. The intent of this breakdown is to provide a logical and convenient way to segment the threat model into manageable chunks that can then be combined back together into a holistic threat model.

The *Device Extracted* `ThreatSendSignal` from Figure 4-8 will trigger the associated `ThreatAcceptEvent` from Figure 4-7, showing that a signal can pass between any level of the mal-activity hierarchy.

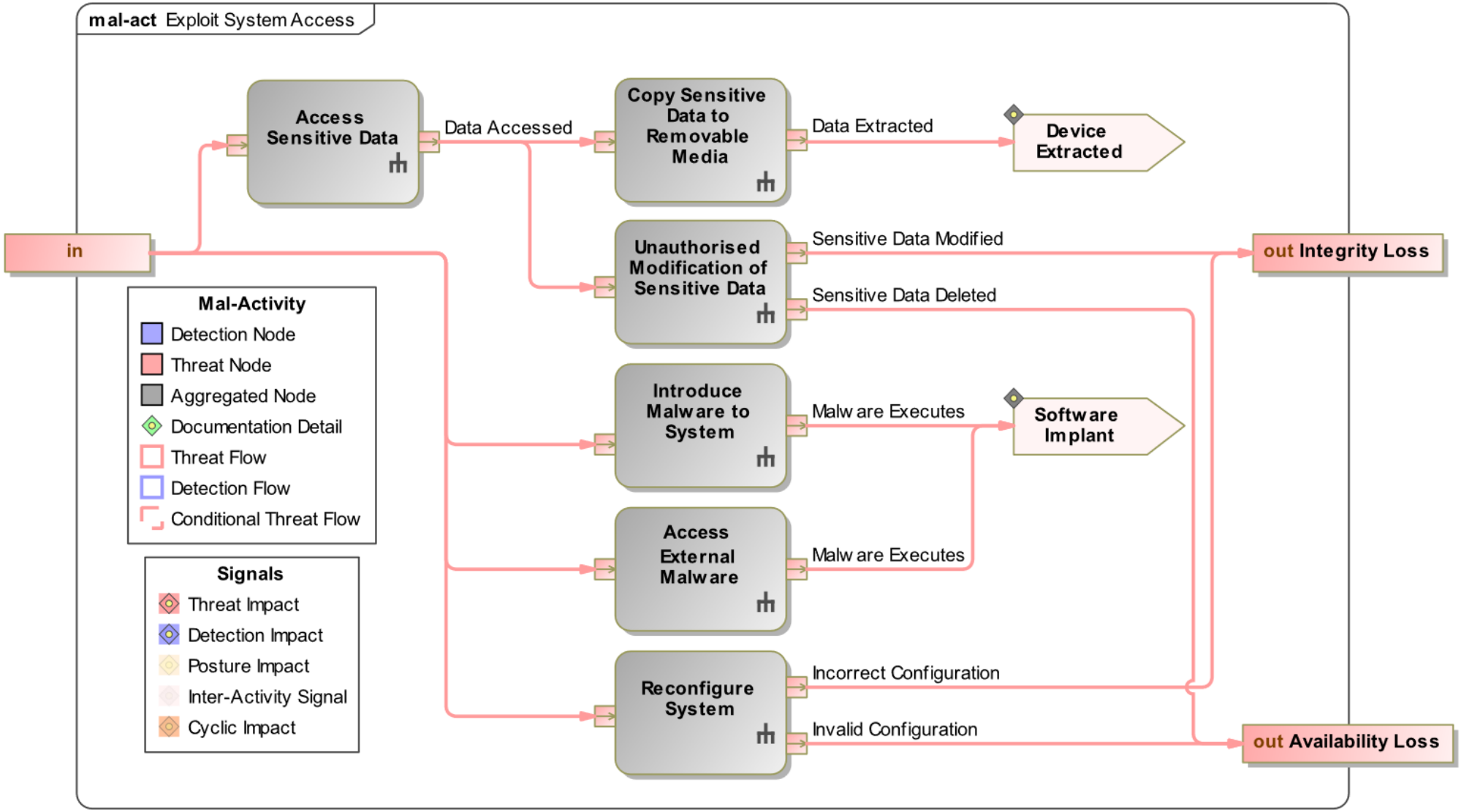


Figure 4-8: *Sample CEMT Mal-Activity Diagram – Exploit System Access*

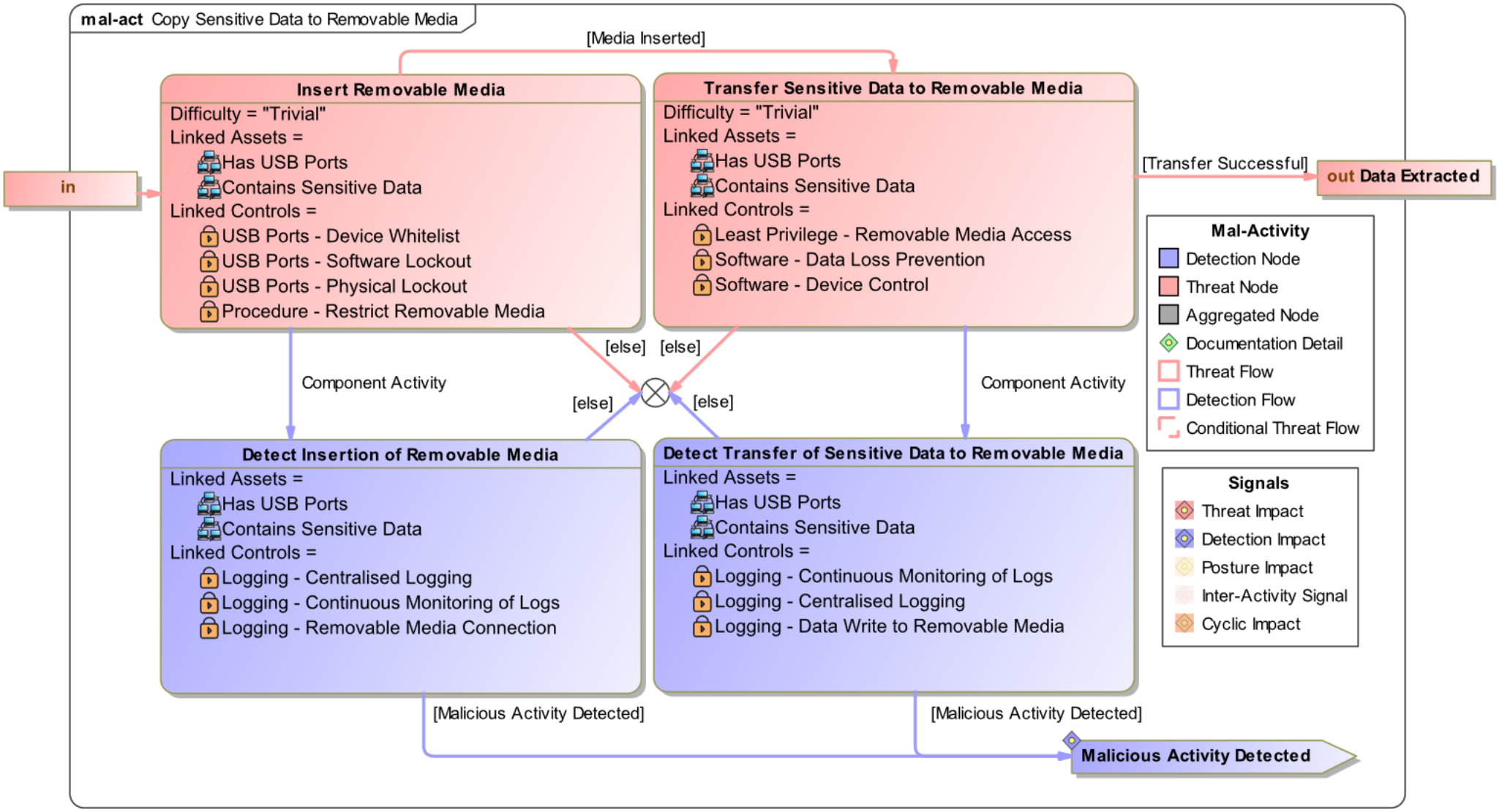


Figure 4-9: *Sample CEMT Mal-Activity Diagram – Copy Sensitive Data*

As the nested hierarchy of mal-activity is built out by the modeller, the bottom level of the hierarchy will eventually be reached and there will no longer be `AggregatedActions` on the *CEMT Mal-Activity Diagram*. Instead, `ThreatActions` and `DetectionActions` will be used to articulate the elemental steps of the threat model, as shown in Figure 4-9 which provides the detail below the *Copy Sensitive Data to Removable Media* `AggregatedAction` from Figure 4-8. As before, the in and out ports on the border of this lower level align with the `ThreatInput` and `ThreatOutput` pins from the higher level drawing. Figure 4-9 also provides an example of a `ThreatEnd` and a `ThreatDetection` being used as termination points for the relevant flows

The lowest level of *CEMT Mal-Activity Diagrams* allows the modeller to start enriching the threat model with additional data for user within the later phases of the CEMT process. Firstly, each `ThreatAction` has a *Difficulty* attribute assigned, which identifies how difficult a particular adversary action is to complete without considering any mitigating security controls. As an example, the *Insert Removable Media* action in Figure 4-9 is marked as *Trivial* because that action does not require any special expertise on behalf of the threat actor, however, an action that represented a complex Electronic-Warfare-enabled cyber attack would require a higher level of adversary expertise to complete even if there were no mitigating controls in place. This *Difficulty* attribute contributes to the likelihood that threat actors of different skill levels will perform the attack in the *Risk Assessment* phase.

Secondly, `AssetGroups` are created and linked into the `ThreatActions` and `DetectionActions` via the `AffectsGroup` relationship. An `AssetGroup` is an attribute of a hypothetical set of components or assets within a system, and the modeller will link those attributes to an action in the threat model when components with the attribute described by the `AssetGroup` would be affected by that action in the threat model. When multiple `AssetGroups` are linked to a single action, a component would need to be a part of all of the linked `AssetGroups` to be considered affected by the action. For example, the *Insert Removable Media* action in Figure 4-9 has two `AssetGroups` linked:

- *Has USB Ports*
- *Contains Sensitive Data*

This combination means that any hypothetical component that had USB ports and contained sensitive data would be affected by this action. Any hypothetical component that did not have both attributes would not be affected by this action because the adversary would not be able to insert a removable media device and copy sensitive data from the component.

Finally, the modeller creates `SecurityControls` and links them to the `ThreatActions` and `DetectionActions` via the `Mitigates` relationship. These security controls are typically drawn from best-practice control sets, such as the Essential Eight Maturity Model [15], the Australian Government Information Security Manual [14] and the MITRE ATT&CK Framework [121] to name a few. These security controls are not intended to be a list of mandatory

controls, or controls that we know are or will be implemented, but rather a list of all feasible controls that could potentially be implemented to mitigate the action to which they are linked. Again, using the *Insert Removable Media* action from Figure 4-9 as an example, there are four linked `SecurityControls`:

- *USB Ports – Device Whitelisting*
- *USB Ports – Software Lockout*
- *USB Ports – Physical Lockout*
- *Procedure – Restrict Removable Media*

These linked `SecurityControls` suggest that one way to mitigate the possibility of an adversary inserting a removable media device (and presumably mounting it in the operating system) would be to have a whitelist of USB devices that would be mounted by the operating system, another way would be to lockout the USB ports in software (such as in the BIOS), another way would be to physically lockout the USB ports with a locking device or glue and another way would be to have a policy that restricts removable media devices. Each of these may have different levels of effectiveness, or may be easier or harder to implement, but that is not important at this phase of the threat model development. The intent here is to simply enumerate all of the controls within the world of feasibility that could sensibly lower the probability that an attacker would be able to perform that action to which they are linked. Essentially, the modeller is distilling the existing knowledge from best-practice control sets and explicitly tracing them to the steps and attacker would take within the threat model.

4.2.3 Attack Trees

Mal-Activities provide a convenient method for building out the threat model, but the nested hierarchy of interconnected flow charts is not a particularly user friendly way for reviewing and understanding the holistic threat model. To address this, the CEMT includes a custom diagram called the *CEMT Attack Tree – Forward* which crawls the `ThreatFlows` in the mal-activity diagrams to identify the various attack paths through the system and presents the attack surface in the form of attack trees. These attack trees are acyclic graphs where each node in the tree represent a node from the mal-activities and each branch of the tree represents a single adversary attack, start to finish.

In any reasonably complex threat model, these attack trees can become unwieldy, with hundreds of thousands of unique branches. To manage this complexity, the default view in the CEMT collapses these trees at the `ThreatSignals` and generates a new tree starting at each `ThreatSignal`. This allows unique segments of the overall attack tree to be presented and reviewed, and the overall path can be traced through via hyperlinks between the different views.

Figure 4-10 shows the forward attack tree starting from the *Physical Incursion* `ThreatStart`. This broadly aligns with the mal-activity diagram shown in Figure 4-7, though the attack tree will show the lowest level of detail available in the model, rather than the intermediate `AggregatedActions` in Figure 4-7. The attack tree shows that there are two paths from a physical attack which both lead

to *Building Access*. Firstly, the attacker might be able to gain access to the building by completing a `ThreatAction`, as shown in the top branch, or they could wait until there is a public open day, at which point the condition described by the `ThreatGuard` on the bottom branch would be satisfied, and the attacker would get direct access to the building. Once they have *Building Access*, the top branch shows that the attacker could perform a `ThreatAction` to gain access to the room. Similarly, Figure 4-11 shows the attack tree starting from the *Insider Threat* `ThreatStart`, which shows that a malicious insider would begin with Room Access by virtue of their status as an insider.

Figure 4-12 displays the attack tree starting from *Room Access*, which shows how an attack would progress from that point and articulates the different options the attacker would have to continue compromising the system once they achieved access to the room. These options include the ability to brute force the login credentials to component in the room, which would provide them *System Access*. Alternatively, they could follow the top branch and remove components from their location in the room, achieving an *Availability Loss* impact, but also reaching the *Device Extracted* signal if they want to continue with an attack to exfiltrate that item from the room.

Figure 4-13 and Figure 4-14 provide examples of an attack tree for the *System Access* and *Device Extracted* signals, respectively, which articulate the various steps a malicious adversary could take from each of those points to continue their attack through to completion.

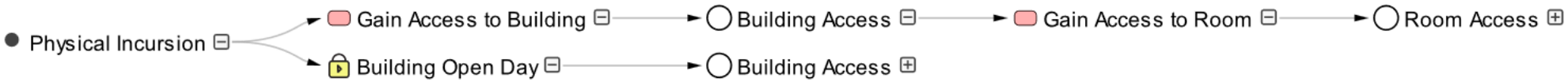


Figure 4-10: *Sample CEMT Forward Attack Tree – Physical Incursion*

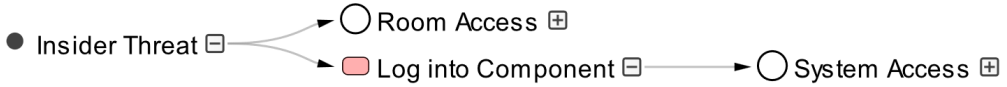


Figure 4-11: *Sample CEMT Forward Attack Tree – Insider Threat*

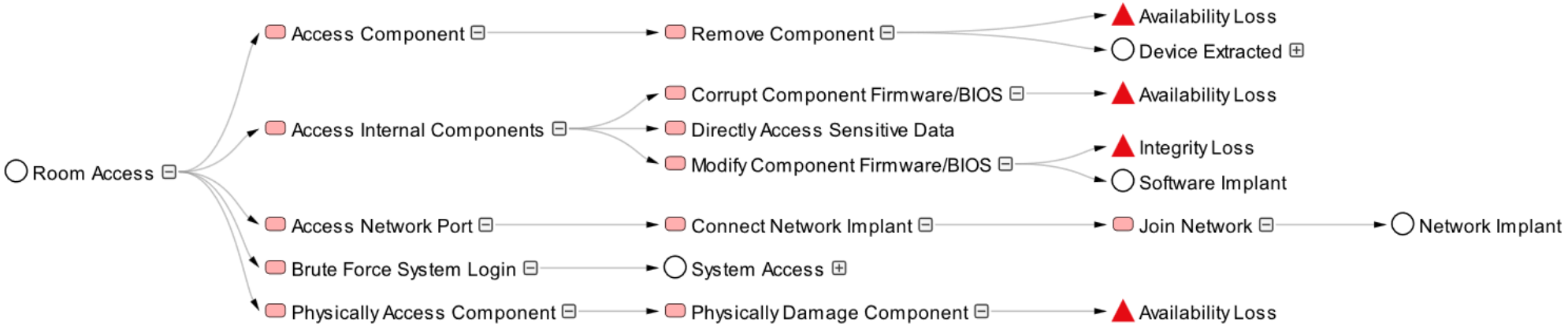


Figure 4-12: *Sample CEMT Forward Attack Tree – Room Access*

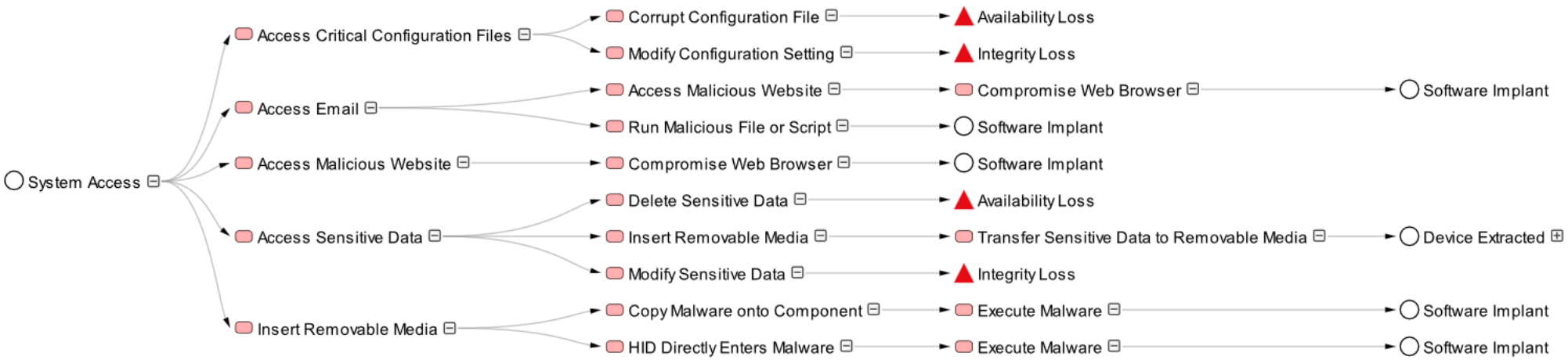


Figure 4-13: *Sample CEMT Forward Attack Tree – System Access*

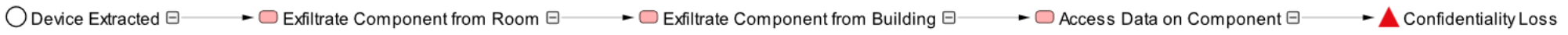


Figure 4-14: *Sample CEMT Forward Attack Tree – Device Extracted*

While the examples provided show the *CEMT Attack Tree – Forward* diagrams, which display all paths leading from a starting point, the CEMT also includes *CEMT Attack Tree – Reverse* diagrams, which display the same relationships but in the opposite direction. This allows a user to select a `ThreatImpactSignal`, such as *Availability Loss* from Figure 4-13, or a `ThreatSignal`, such as *Room Access* from Figure 4-12, and the CEMT would display all of the different attack paths that could lead to that outcome. This allows a user to see the different methods by which an attacker might attempt to achieve an undesired outcome.

These attack trees, by themselves, can start providing insights into the security of a system being evaluated. Nodes that are common across multiple branches of the attack tree are likely to more efficient targets for implementing security mitigations than nodes that only appear on a handful of paths. Additionally, when the number of nodes in branch on the attack tree is particularly low, this can indicate a short attack path that requires attention as there is a lack of natural defence-in-depth.

While this sort of assessment can be performed manually, the CEMT also includes automated assessments to identify short branches, and nodes that exist on multiple branches. The CEMT can prioritise `SecurityControls` based on how much that control participates in the overall threat model – that is, it counts how many branches of the attack tree each `SecurityControl` is mitigating and then normalises that number against the `SecurityControl` with the highest participation count to produce a relative importance for each `SecurityControl`.

4.2.4 Threat Sets

The *Threat Modelling* phase finishes once the attack trees have been developed, and it is important to recognise that the components in the system have not been required to complete this phase of the CEMT process. The threat model at this stage is largely agnostic to the components within the system and does not rely on any implementation level details being understood for the system. While there are some specific constraints implied by the sequence of events articulated in the model, the attack trees are still relatively generic. This generalisation leads to two desirable properties of the threat model.

Firstly, the threat model could be completed during an early conceptual design phase, where the functions of a system may be understood, but the detailed design has not yet been completed. This allows the security evaluation to begin early in the development lifecycle, where the cost to implement changes to address any security concerns is lower than later stages of the lifecycle.

Secondly, a generic model can be readily reused across similar system in the future. The fact that the output of the *Threat Modelling* phase is agnostic to the specific components in the system means that reusable threat sets can be generated and centrally managed within an organisation so that they can be rapidly reapplied on similar systems in the future. These generic threat sets can be distributed as shared libraries within the model-based systems engineering tool, ensuring consistency across systems and facilitating centralised updates to those models over time that are automatically flowed out to each system that

utilises those centralised libraries. Where specific threats exist in a particular system being assessed, these can be added alongside the generic threat set library to ensure that each system under evaluation receives a contextualised threat model without requiring models to reinvent the wheel during every evaluation.

`SecurityConstraints` can also be incorporated into these centralised threat sets. As described on page 113, a `SecurityConstraint` is a special type of `SecurityControl` which can be linked directly to an `AssetGroup` via the `AppliesGroup` relationship in addition to the normal link to a `ThreatModelAction` via the `Mitigates` relationship. This allows for mandatory security compliance requirements to be incorporated into a CEMT assessment even if they do not directly address a threat articulated in the threat model. While this isn't described here in detail, because ideally only mitigations that are directly related to a threat would be part of a security risk assessment, the `SecurityConstraint` stereotype is included in the CEMT to address the reality that compliance requirements are usually unavoidable in many industries.

The CEMT does help with reusability and efficiency of `SecurityConstraint` based assessments, as they can be linked to `AssetGroups` as part of a sharable threat set, which can help save time by streamlining the allocation of these compliance requirements to relevant components during the *Threat Allocation* phase.

4.3 THREAT ALLOCATION

The *Threat Allocation* phase involves taking the threat sets developed in the *Threat Modelling* phase and allocating those threat sets to the components within the system being evaluated. This requires the creation of an asset hierarchy for the system, assigning those assets into the `AssetGroups` defined in the threat sets, generating a threat-informed control baseline for each asset and then assessing a control implementation state for each control on each asset.

4.3.1 Asset Definition

The first step in the *Threat Allocation* phase is the definition of the asset hierarchy. An asset is defined as any component within the system breakdown structure that is security relevant – that is, if it could be the target of any part of an adversary attack or could implement security mitigations against those attacks. By this definition, an asset doesn't have to be part of the system but could be an external component that is relevant to the security of the system itself, such as the physical building in which the system is located or an external system that is interfaced into the system being evaluated.

Like most component definition problems within systems engineering, the answer to how detailed the asset hierarchy should be is going to be different for each assessment. This will usually be based on what level of the breakdown you are able to influence with your design as there will be little value in trying to further decompose a component that is already considered a 'black-box' in your design.

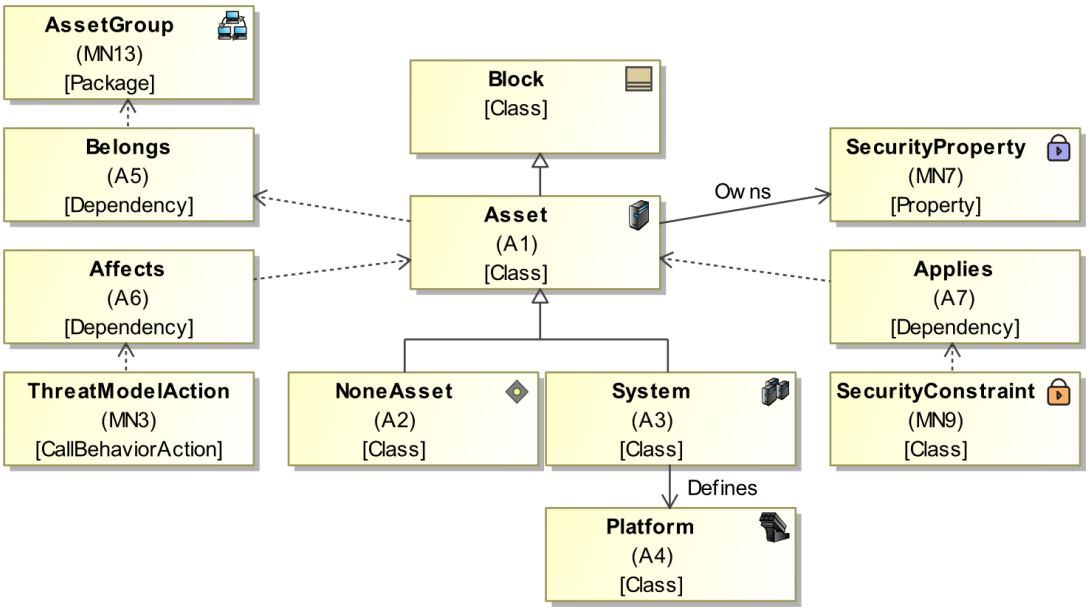


Figure 4-15: *CEMT Asset Taxonomy*

Figure 4-15 shows the taxonomy of stereotypes that the CEMT includes for defining the assets and their relationships to other CEMT objects:

A1. `Asset` – This stereotype inherits from the SysML `Block` stereotype and is used to represent any component that could be the target of any part of an adversary attack or could implement security mitigations against those attacks. This stereotype can be linked to an `AssetGroup` via the `Belongs` dependency, can be linked from a `ThreatModelAction` via the `Affects` dependency and can be linked from a `SecurityConstraint` via the `Applies` dependency.

A2. `NoneAsset` – This stereotype inherits from the `Asset` stereotype and is used to create a special `Asset` that represents the lack of any relevant `Assets`. This `NoneAsset` is linked to any `SecurityConstraints` that are not applicable to any `Assets` in the

system to differentiate a control that has been assessed as having no applicable assets from one which has not yet been assessed.

A3. `System` – This stereotype inherits from the `Asset` stereotype and is used to identify the system under evaluation. Typically, there is only one `System` in each CEMT model.

A4. `Platform` – This stereotype inherits from the `Asset` stereotype and is used to enumerate the different instances of the system under evaluation. This allows for multiple platforms to be assessed, with their own risk assessments, using the common threat model and asset hierarchy.

A5. `Belongs` – This stereotype inherits from the built-in `Dependency` metaclass and is used as a relation between an `Asset` and an `AssetGroup` to show that the `Asset` has the attributes described by the `AssetGroup`.

A6. `Affects` – This stereotype inherits from the built-in `Dependency` metaclass and is used as a relation between an `Asset` and an `ThreatModelAction` to show that the `Asset` is affected by the `ThreatModelAction`.

A7. `Applies` – This stereotype inherits from the built-in `Dependency` metaclass and is used as a relation between an `Asset` and an `SecurityConstraint` to show that the `SecurityConstraint` applies to the `AssetGroup`.

These stereotypes are used by the modeller to produce *CEMT Asset Definition Diagrams*, which are custom diagrams within the CEMT that are based on built-in *SysML Block Definition Diagrams*. An example of a *CEMT Asset Definition Diagram* is shown in Figure 4-16.

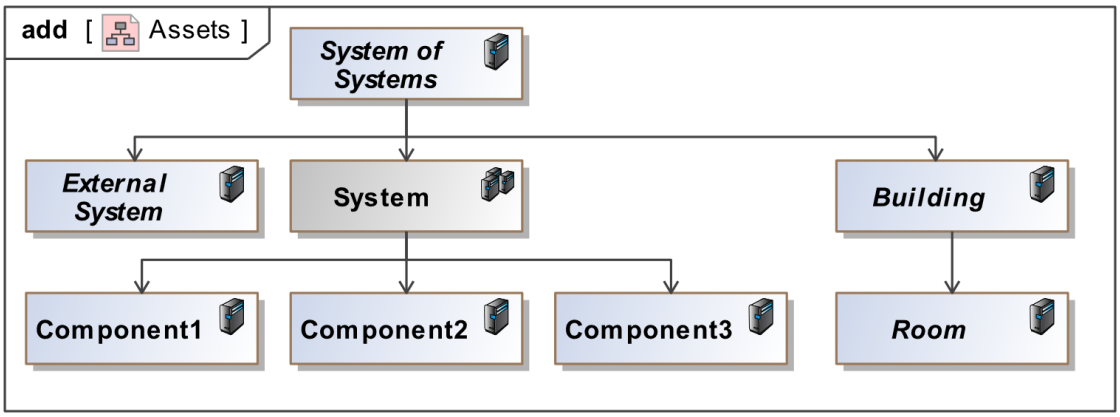


Figure 4-16: *Sample CEMT Asset Definition Diagram*

The *CEMT Asset Definition Diagram* is centred around the `Asset` and `System` objects. In Figure 4-3, *System* is an example of a `System` object, while all of the other objects are `Asset` objects. As shown, `Assets` can include both components within the system, as well as components outside of the system. These objects are all linked together with the built-in Directed Association dependency to build a hierarchy that reflects the system breakdown structure.

The second step in the *Threat Allocation* phase requires the modeller to allocate all of the `Assets` in the asset hierarchy to `AssetGroups` from the threat sets developed during the earlier phases. This allocation is based on the properties of the asset and the attributes described by asset group. As an example, consider two assets:

- *Asset-1* – A workstation PC that processes and stores an organisation's sensitive data and has two UBS ports
- *Asset-2* – A standalone personal laptop that does not store or process an organisation's sensitive data, has one USB port and is used for personal internet browsing during lunch breaks

Our threat set contains three asset groups:

- *Has USB Ports* – which covers all devices with USB ports
- *Contains Sensitive Data* – which covers all devices that store and/or process sensitive data
- *Portable Device* – which covers all devices that do not have a consistent physical location when operating

In this example, *Asset-1* would be linked to *Has USB Ports* and *Contains Sensitive Data* while *Asset-2* would be linked to *Has USB Ports* and *Portable Device*.

As the `AssetGroups` are linked to the `ThreatModelActions` by the `AffectsGroup` relationship in the threat set, this allocation of `Assets` to `AssetGroups` creates an indirect relationship between `Assets` in the asset hierarchy and `ThreatModelActions` in the threat set. Based on this indirect relationship, the CEMT will create a direct `Affects` relationship between each `Asset` and the related `ThreatModelActions` from the threat set.

Furthermore, because the `ThreatModelActions` are linked to `SecurityControls` by the `Mitigates` relationship, there is an indirect relationship between each `Asset` and the `SecurityControls` that the threat model has assessed as relevant to that asset. Based on this relationship, the CEMT will create a `SecurityProperty`, owned by the `Asset` and typed by the `SecurityControl`, which represents an instantiation of each relevant control onto each asset.

Ultimately, this generates a set of security mitigations for each asset that is derived from the threat model. This is the threat-informed security control baseline for the system under evaluation, which is a key output of the *Threat Allocation* phase.

4.3.2 Control Implementation

Once a threat-informed control baseline has been derived, the implementation state of those controls must be determined and entered into the threat model. While an exhaustive audit of all controls would be desirable, on large systems with a lot of assets, it may be acceptable to prioritise which controls are audited based on the relative priority that the CEMT can calculate for each `SecurityProperty` based on the importance of the `SecurityControl` and the criticality of the `Asset`. Ultimately, the exhaustiveness of the audit will be determined on a system-by-system basis. The minimum requirement of the CEMT is that all security controls which are referenced in the following *Risk Assessment* phase must have an implementation state set.

Each `SecurityProperty` contains three different implementation states that can exist side-by-side in the threat model:

- *Designed* – which represents the implementation state in the current design baseline
- *Verified* – which represents an implementation state which has been confirmed via audit, inspection or test
- *Proposed* – which represents an implementation state of an alternate design baseline, such as an intended change to the system

These three implementation states are used to produce different risk assessments during the subsequent phases of the CEMT process. A risk assessment based on the *Verified* implementation state of controls provides an assessment with greater assurance over the actual implementation baseline. A risk assessment based on the *Designed* implementation state of controls provides a less assured, but often more current risk assessment. A risk assessment based on the *Proposed* implementation state projects an expected residual risk if additional controls beyond the current baseline were to be implemented, supporting cost-benefit analyses for those additional control. Risk assessments for each of these implementation states can be stored within the threat model simultaneously.

4.4 RISK ASSESSMENT

The *Risk Assessment* phase involves the identification of critical branches of the attack tree for quantitative analysis, using numerical simulation to complete a risk assessment of the key cyber risks within the system under evaluation. These risks are categorised according to the adversary's mission and are described using parametric diagrams that incorporate threat actor capability, threat actor intent and control effectiveness.

Within the CEMT, a security risk represents a single attack from inception through to an impact on the system's confidentiality, integrity or availability being realised. Each full branch of the attack tree, from the initial starting point of a misuse case through to an impact signal, is a potential security risk that can be analysed during the *Risk Assessment* phase.

4.4.1 Risk Modules

Risk modules are critical building blocks for performing risk assessments using the CEMT and represent a string of sequential threat nodes that form a subset of a branch of the attack tree. These risk modules are used to calculate the likelihood that an attacker will progress through the sequence of threat nodes represented by the risk module. Multiple risk modules can be chained together into different combinations, provided the overall combination is a valid branch of the attack tree. The combination of risk modules can be used to calculate the likelihood that an attacker would be able to perform the complete attack.

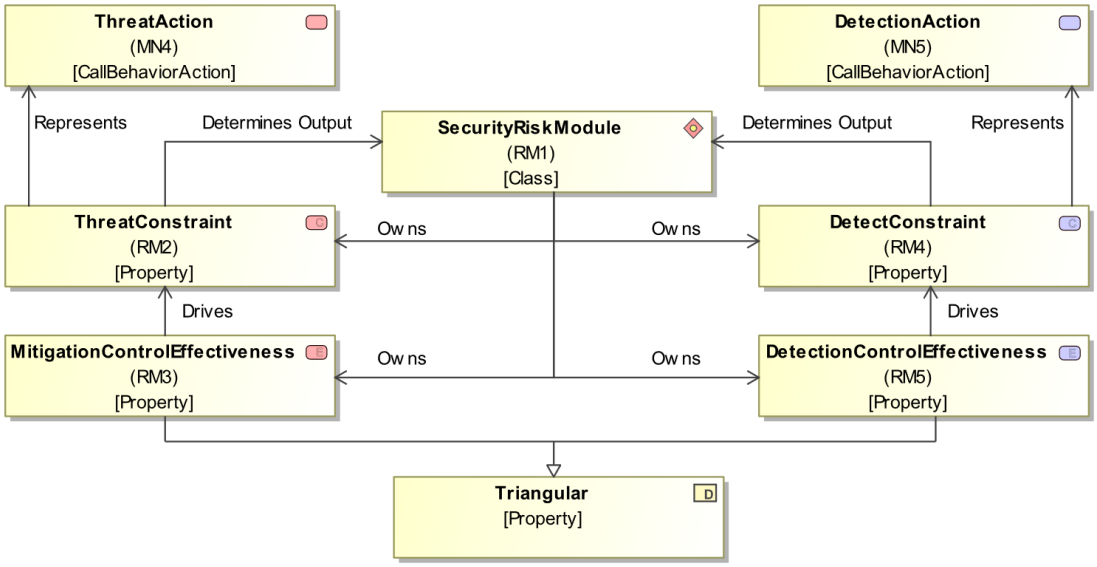


Figure 4-17: *CEMT Security Risk Module Taxonomy*

Figure 4-17 shows the taxonomy of stereotypes that the CEMT includes for defining the security risk modules:

RM1. `SecurityRiskModule` – This stereotype inherits from the SysML `Block` stereotype and is used to contain the properties and constraints that calculate the likelihood a threat actor will be able to perform each node in the attack tree.

RM2. `ThreatConstraint` – This stereotype inherits from the SysML `ConstraintProperty` stereotype, represents a threat node in the attack tree and converts input values into a likelihood that the attacker will reach a particular point in the attack tree.

RM3. `MitigationControlEffectiveness` – This stereotype inherits from the built-in `Triangular` stereotype and provides a three-point estimate for the effectiveness of the controls linked to a particular threat

node, which is defined as the percentage of attempted attacks that reached this point that would be prevented by those controls.

RM4. `DetectConstraint` – This stereotype inherits from the SysML `ConstraintProperty` stereotype, represents a detection node connected to a threat node in the attack tree and converts input values into a likelihood that the attacker will be detected by the time they reach a particular point in the attack tree.

RM5. `DetectionControlEffectiveness` – This stereotype inherits from the built-in `Triangular` stereotype and provides a three-point estimate for the effectiveness of the controls linked to a particular detection node, which is defined as the percentage of attempted attacks that reach this point that would be detected by those controls.

These stereotypes are used by the modeller to produce a *CEMT Parametric Risk Diagram*, which is a custom diagram within the CEMT that is based on the built-in *SysML Parametric Diagram*, that defines the internal structure of a `SecurityRiskModule`. An example of a *CEMT Parametric Risk Diagram* is split over two figures to aid readability, with the top half of the diagram shown in Figure 4-18 and the bottom half of the diagram shown in Figure 4-19.

The modeller creates a risk module by selecting nodes from the attack tree, along with a set of assets that they want to include in the risk module, and then the CEMT will create the `SecurityRiskModule` and build a *CEMT Parametric Risk Diagram* centred on those selected objects.

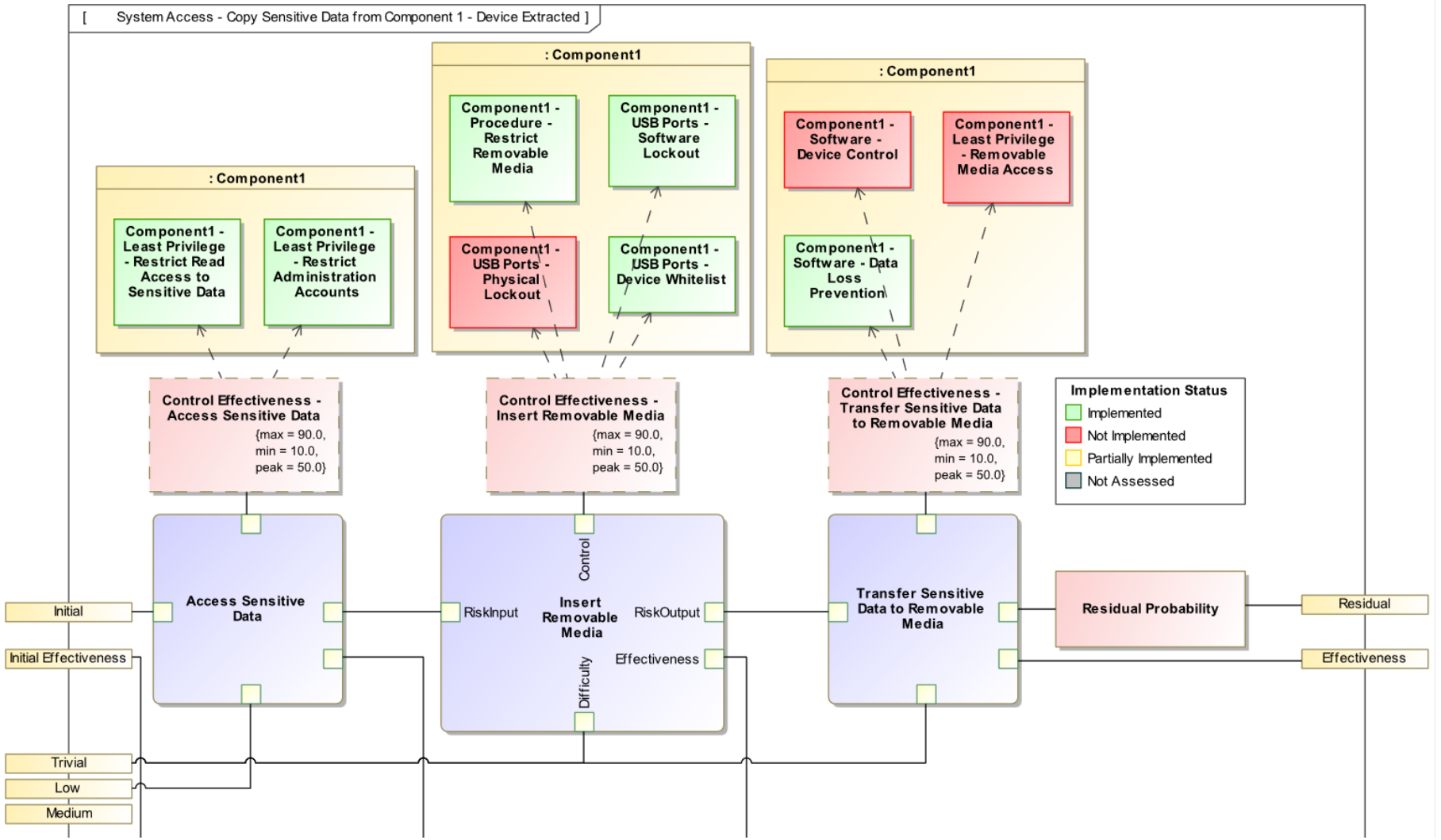


Figure 4-18: *Sample CEMT Parametric Risk Diagram – Risk Module – Threat Actions*

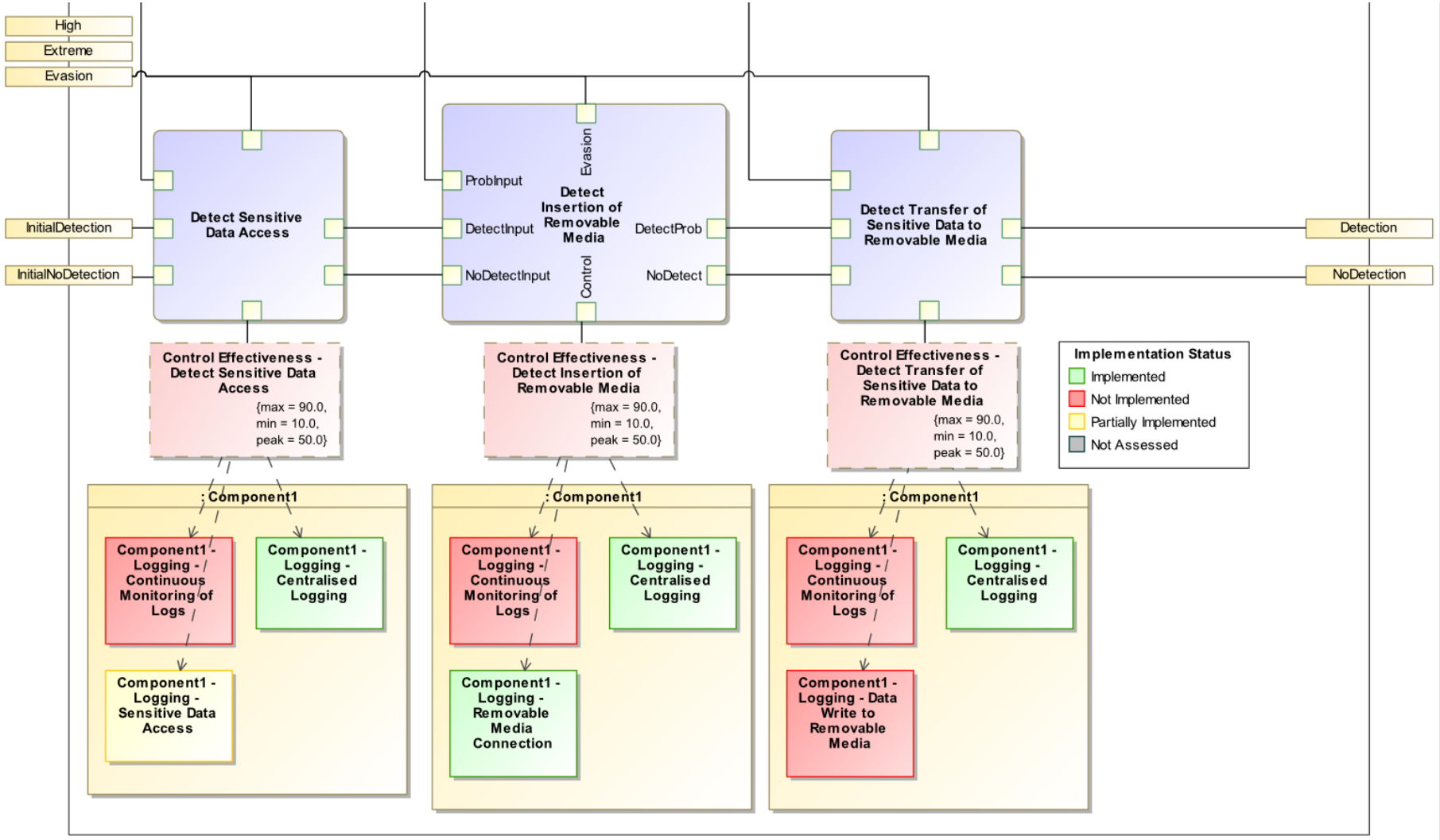


Figure 4-19: *Sample CEMT Parametric Risk Diagram – Risk Module – Detection Actions*

In the case of the sample diagram in Figure 4-18 and Figure 4-19, the modeller would have selected the *Component1* asset and the following `ThreatActions`:

- *Access Sensitive Data*
- *Insert Removable Media*
- *Transfer Sensitive Data to Removable Media*

The CEMT then created a `ThreatConstraint` for each of those `ThreatActions` with an associated `MitigationControlEffectiveness` and linked the `SecurityProperties` owned by the *Component1* asset that are related to the `ThreatAction` to create the diagram in Figure 4-18. Each of the `ThreatConstraints` are also connected to a parameter on the border of the diagram based on the *Difficulty* of the `ThreatAction` is represents.

`DetectConstraints` for the `DetectionActions` associated with the selected `ThreatActions` are then created by the CEMT, linked to a `DetectionControlEffectiveness` which is then linked to the relevant `SecurityProperties` associated with the `DetectionAction` to create the diagram in Figure 4-19. `SecurityProperties` in the *CEMT Parametric Risk Diagrams* are colour-coded based on the implementation status determined during the *Threat Allocation* phase.

For each node in the diagram, the modeller needs to determine how effective the implemented controls are at each preventing the adversary from progressing with the attack, and how effective the implemented controls are at detecting the

adversary as they perform the attack. This effectiveness is expressed as the percentage of attempts that would be either stopped or detected by the controls, and the modeller determines a threat-point estimate with *min*, *max* and *peak* values to create a triangular distribution to incorporate any uncertainty in the effectiveness determination.

The *Insert Removable Media* node in Figure 4-18 has been modified to display the input and output parameters of the `ThreatConstraint`:

- *RiskInput* – input parameter for the probability that the adversary has reached this node, expressed as a number between 0 and 100.
- *Control* – input parameter for the effectiveness of controls linked to this node, representing the percentage of attacks that would be prevented by the implemented controls, expressed as a number between 0 and 100.
- *Difficulty* – input parameter for the probability that an adversary would be able to complete a node with the difficulty of the current node, which is dependent on the capability of the threat actor and is expressed as number between 0 and 100.
- *Effectiveness* – output parameter that combines Control and Difficulty to calculate the percentage of attempts at this node that would be successfully completed, expressed as a number between 0 and 100. This output feeds into the *ProbInput* parameter of the `DetectConstraint` associated with the next `ThreatConstraint` in the path. Equation (1) shows how *Effectiveness* is calculated.

- *RiskOutput* – output parameter that combines *Effectiveness* and *RiskInput* to calculate the probability that the adversary has passed this node, expressed as a number between 0 and 100. This output feeds into the *RiskInput* parameter of the next `ThreatConstraint` in the path. Equation (2) shows how *RiskOutput* is calculated.

The equations that calculate the outputs of the `ThreatConstraint` are shown below:

$$\frac{Effectiveness}{100} = \left(1 - \frac{Control}{100}\right) \times \frac{Difficulty}{100} \quad (1)$$

$$\frac{RiskOutput}{100} = \frac{Effectiveness}{100} \times \frac{RiskInput}{100} \quad (2)$$

The *Detect Insertion of Removable Media* node in Figure 4-19 has been modified to display the input and output parameters of the `DetectConstraint`:

- *ProbInput* – input parameter for the probability that the attacker would not be able to progress past the previous node in the attack path, expressed as a number between 0 and 100.
- *DetectInput* – input parameter for the probability that the attack has been detected at any point before this node, expressed as a number between 0 and 100.
- *NoDetectInput* – input parameter for the probability that the attacker reached the previous detection node without being detected at any stage

in the path to that point and was not detected by the previous detection node, expressed as a number between 0 and 100.

- *Control* – input parameter for the effectiveness of controls linked to this node, representing the percentage of attacks that would be detected by the implemented controls, expressed as a number between 0 and 100.
- *Evasion* – input parameter for the probability that an adversary would be able to evade detection at the current node, which is dependent on the capability of the threat actor and is expressed as number between 0 and 100.
- *DetectProb* – output parameter that combines all input parameters to calculate the probability that the attack has been detected at any point in the path up to and including this node, expressed as a number between 0 and 100. This output feeds into the *DetectInput* parameter of the next `DetectConstraint` in the path. Equation (4) shows how *DetectProb* is calculated.
- *NoDetect* – output parameter that combines *ProbInput*, *DetectInput*, *NoDetectInput*, *Control* and *Evasion* to calculate the probability that the adversary reached this node without being previously detected and wasn't detected by this node, expressed as a number between 0 and 100. This output feeds into the *NoDetectInput* parameter of the next `DetectConstraint` in the path. Equation (5) shows how *NoDetect* is calculated.

The equations that calculate the outputs of the `DetectConstraint` are shown below:

$$P_D = \frac{Control}{100} \times \left(1 - \frac{Evasion}{100}\right) \quad (3)$$

$$\frac{DetectProb}{100} = \left(\frac{NoDetectInput}{100} \times \frac{ProbInput}{100} \times P_D\right) + \frac{DetectInput}{100} \quad (4)$$

$$\frac{NoDetectProb}{100} = \frac{NoDetectInput}{100} \times \frac{ProbInput}{100} \times (1 - P_D) \quad (5)$$

Ultimately, these constraints keep a running total of the probability that an attacker has reached a certain point in the attack tree, as well as the probability that the attacker has been detected while performing those actions.

It is important to understand that the calculation treats each node as an independent event, but there are many occasions where these events might not be truly independent. As an example, if an attacker was not stopped by a set of controls earlier in an attack path, those same controls are unlikely to stop the same attacker if they were mitigating another node later in the same attack path.

This dependence needs to be handled by the modeller with the control effectiveness values – if the same control appears twice in the same risk module, the control effectiveness should be lower for the second instance to account for the fact that the controls are not effective if they can be bypassed by the attacker.

4.4.2 Security Risks

Security risks are a combination of risk modules which together represent a full branch of the attack tree, from the initial node of the misuse case through to the undesirable impact being realised. The security risk also includes several input variables that drive the chain of risk modules, and output variables that track the overall likelihood that the risk will be realised and the likelihood that an attack would be detected.

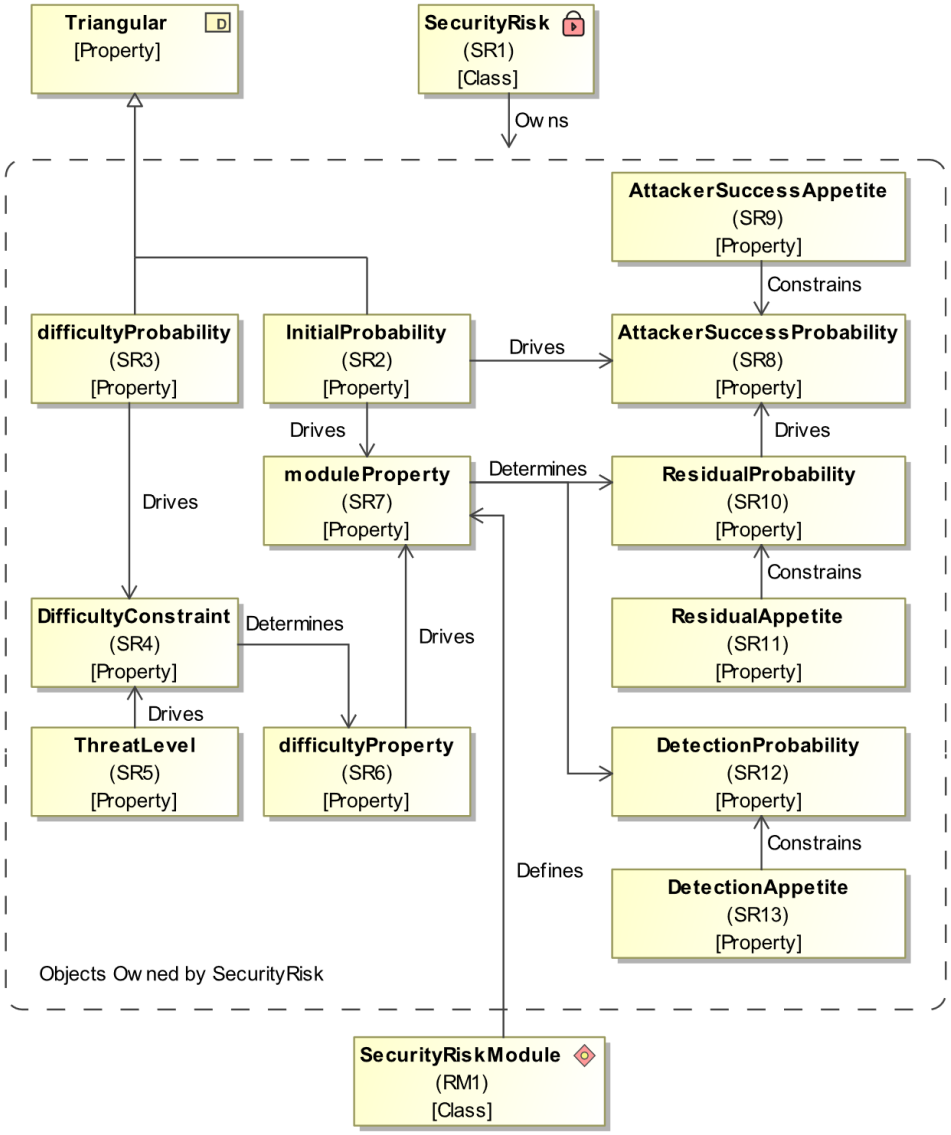


Figure 4-20: *CEMT Security Risk Taxonomy*

Figure 4-20 shows the taxonomy of stereotypes that the CEMT includes for defining the security risks:

SR1. `SecurityRisk` – This stereotype inherits from the SysML `Block` stereotype and is used to contain the properties and constraints that calculate the likelihood a threat actor will be able to perform an entire branch of the attack tree.

SR2. `InitialProbability` – The stereotype inherits from the built-in `Triangular` stereotype and provides a three-point estimate for the probability that the security risk will be attempted by an adversary. This stereotype is an input to the first `moduleProperty` and the `AttackerSuccessProbability` in the security risk.

SR3. `difficultyProbability` – This stereotype inherits from the built-in `Triangular` stereotype and provides a three-point estimate that quantifies qualitative likelihood bands that describe how likely a threat actor is to perform a threat action.

SR4. `DifficultyConstraint` – This stereotype inherits from the SysML `ConstraintProperty` stereotype and uses a lookup table to that sets the `difficultyProperty` for each difficulty level based on the qualitative likelihood bands taken from the `difficultyProbability` inputs and the current `ThreatLevel`.

SR5. `ThreatLevel` – This stereotype inherits from the SysML ValueProperty stereotype and represents the capability level of the threat actor performing the attack described by the security risk. The

`ThreatLevel` can be set to either *Nation State*, *Professional*, *Intermediate* or *Novice*.

SR6. `difficultyProperty` – This stereotype inherits from the SysML `ValueProperty` stereotype, represents each of the possible difficulty levels for the nodes of the attack path and is set to the probability that current threat actor will be able complete `ThreatActions` that are set to the same difficulty as the `difficultyProperty`.

SR7. `moduleProperty` – This stereotype inherits from the SysML `PartProperty` stereotype and is defined by a `SecurityRiskModule`. These properties are chained together to describe a complete branch of the attack tree and are then connected to the relevant input and output variables in the security risk.

SR8. `AttackerSuccessProbability` – This stereotype inherits from the SysML `ValueProperty` stereotype, is one of the three primary outputs of the security risk and contains the likelihood that an attack will be successful if it is attempted by the attacker.

SR9. `AttackerSuccessAppetite` – This stereotype inherits from the SysML `ValueProperty` stereotype and provides a threshold value for what level of attacker success rate is acceptable to decision makers.

SR10. `ResidualProbability` – This stereotype inherits from the SysML `ValueProperty` stereotype, is one of the three primary outputs of the security risk and contains the likelihood that an attack will be attempted and completed by the attacker.

SR11. `ResidualAppetite` – This stereotype inherits from the SysML `ValueProperty` stereotype and provides a threshold value for what level of risk likelihood is acceptable to decision makers.

SR12. `DetectionProbability` – This stereotype inherits from the SysML `ValueProperty` stereotype, is one of the three primary outputs of the security risk and contains the likelihood that an attack will be detected if it is attempted by the attacker.

SR13. `DetectionAppetite` – This stereotype inherits from the SysML `ValueProperty` stereotype and provides a threshold value for what level of detection likelihood is acceptable to decision makers.

These stereotypes are used by the modeller to produce a *CEMT Parametric Risk Diagram*, which is a custom diagram within the CEMT that is based on the built-in *SysML Parametric Diagram*, that defines the internal structure of a `SecurityRisk`. An example of a *CEMT Parametric Risk Diagram* that defines the security risk is shown in Figure 4-21.

The modeller creates a security risk by selecting multiple risk modules that form a full branch of the attack tree when connected end-to-end, and then the CEMT will create the `SecurityRisk` and build a *CEMT Parametric Risk Diagram* centred on those selected risk modules.

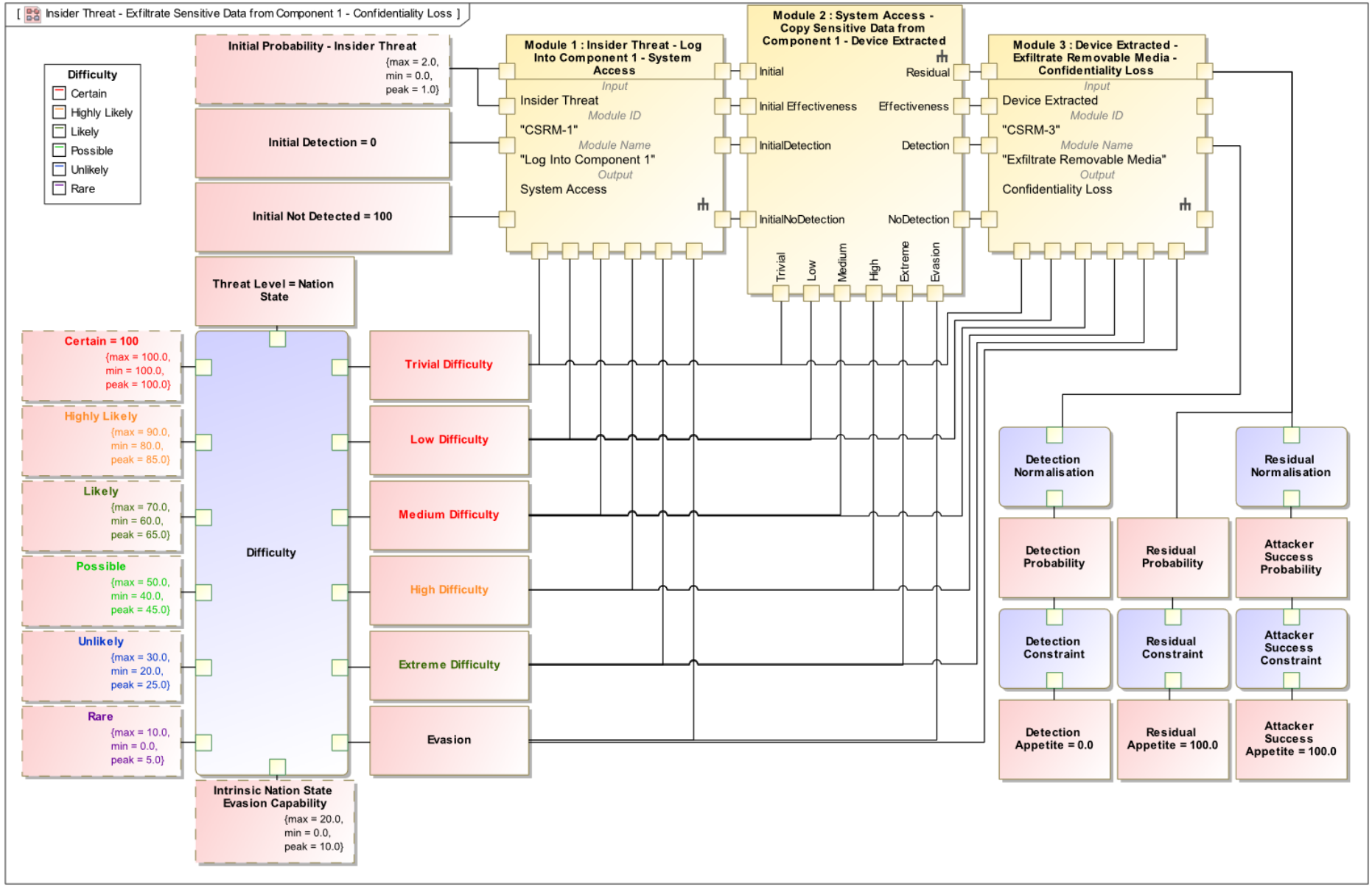


Figure 4-21: *Sample CEMT Parametric Risk Diagram – Security Risk*

In the case of the sample diagram in Figure 4-21, the modeller would have selected the following `SecurityRiskModules`:

- *Insider Threat – Log Into Component1 – System Access*
- *System Access – Copy Sensitive Data from Component1 – Device Extracted*
- *Device Extracted – Exfiltrate Removable Media – Confidentiality Loss*

The CEMT then created a `moduleProperty` for each of those `SecurityRiskModules` and joined their inputs and outputs together to form a chain of risk modules. The *System Access – Copy Sensitive Data from Component1 – Device Extracted* module in Figure 4-21 has been modified to display the input and output ports of the `SecurityRiskModule`, which align with the ports on the external border of the internal diagram of the module shown in Figure 4-18 and Figure 4-19.

The difficulty ports (*Trivial*, *Low*, *Medium*, *High*, *Extreme* and *Evasion*) of each risk module are connected to a corresponding `difficultyProperty`. A `DifficultyConstraint` (*Difficulty* in Figure 4-21) is created, along with the `ThreatLevel` and the `difficultyProbabilities` that are the inputs to that constraint.

The CEMT creates an `InitialProbability` and the *Initial Detection* and *Initial Not Detection* values and connects them to the input ports of the first risk module as shown in Figure 4-21. The output values for residual probability, detection

probability and attacker success rate are also created and linked to the output ports of the final risk module.

Once the diagram has been created, the modeller needs to select a threat level that they would like to assess and set a three-point estimate for the initial probability that reflects their assessment of the likelihood that the attack described by the security risk will be performed by an adversary. These values are primarily derived from threat intelligence sources, with the threat level reflecting the threat capability and the initial probability reflecting threat intent. With these values determined and in place, the security risk is ready to be numerically simulated within the CEMT.

4.4.3 Simulation

Following the creation of the security risk, the next step in the *Risk Assessment* phase is to leverage Monte Carlo simulations to perform a quantitative risk assessment of the security risks.

The CEMT simulates the security risk by first randomly selecting an input value from within the triangular probability distributions that the modeller defined for all input variables:

- Initial Probability
- Difficulty Probabilities
- Control Effectiveness (for the risk modules referenced by the risk)

The CEMT then uses those selected values as inputs to the various constraint properties which eventually produces the output values for the risk:

- Residual Probability
- Detection Probability
- Attacker Success Probability

To account for the uncertainty of the input values, the CEMT completes the Monte Carlo simulation by repeating this process and tracking the results to produce a probability distribution for the output values. The resultant distribution of the output values is displayed on a histogram, as shown in Figure 4-22.

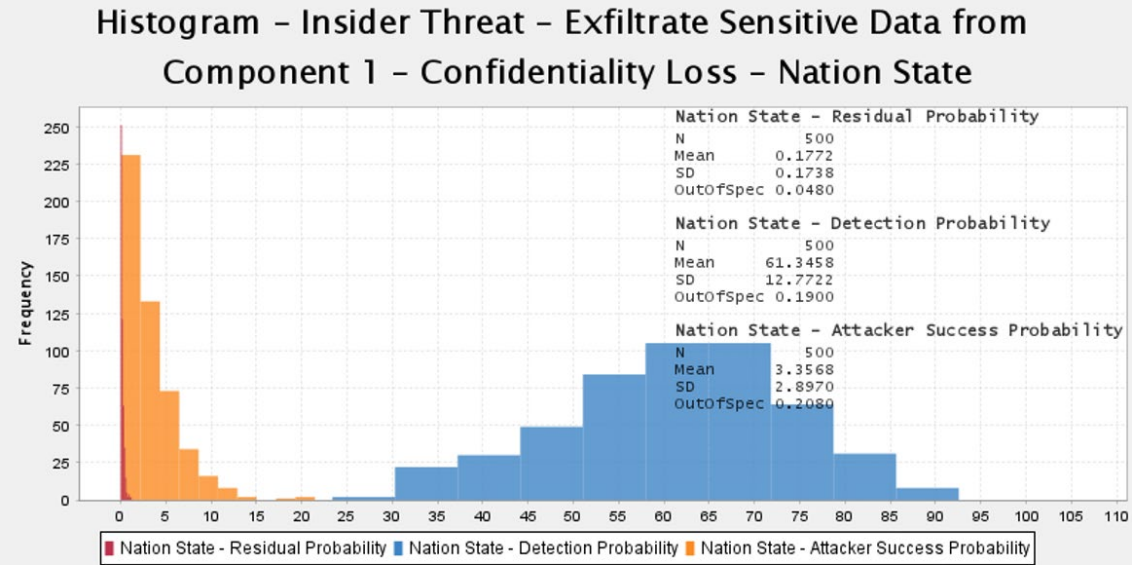


Figure 4-22: *CEMT Simulation Histogram*

For each of the output values, the histogram includes the following data:

- *N* – The number of runs in the Monte Carlo simulation
- *Mean* – The average value of the output variable across the runs

- *SD* – The standard deviation of the output variable across the runs
- *OutOfSpec* – The percentage of runs where the output value was beyond the threshold value specified in the associated appetite property, expressed as a number between 0 and 1

For the sample simulation shown in Figure 4-22, there were 500 runs simulated for this risk. The residual probability average was 0.1772, which means the probability of the attack being attempted and succeeding was 0.1772%. The residual probability *OutOfSpec* was 0.0480, which means that on 4.8% of the 500 runs the residual probability exceeded the residual probability appetite. The detection probability average was 61.3458, which means the probability of the attack being detected if it was attempted was 61.3458%. The detection probability *OutOfSpec* was 0.1900, which means that on 19% of the 500 runs the detection probability was lower than the detection probability appetite. The attacker success probability average was 3.3568, which means the probability of the attack being detected if it was attempted was 3.3568%. The attacker success probability *OutOfSpec* was 0.2080, which means that on 20.8% of the 500 runs the attacker success probability exceeded the attacker success probability appetite.

Ultimately, the CEMT simulation provides a quantitative assessment of the likelihood of the security risk, based on an initial likelihood of the attack being attempted, the capability of the threat actor and the effectiveness of controls within the system being evaluated. The distribution on the produced histogram

provides insight into the uncertain around those quantitative values, as well as the probability that the assessed risk is beyond the defined risk appetite thresholds.

4.4.4 Risk Categorisation and Analysis

Once the security risks have been created and assessed, those risks are categorised to assist with contextualising the risks for decision makers. This categorisation is performed across two metrics. Firstly, the risks are categorised into groupings based on the type of attack being performed by the adversary and secondly, the risks are categorised into qualitative risk ratings.

Additionally, there are methods that a modeller can use to streamline the simulation process by using aggregated objects that perform multiple simulations in parallel. This allows the modeller to perform the quantitative analysis of the system more rapidly to get better coverage across different threat actors and different threat environments in which the system under evaluation might be operated.

This section outlines how the CEMT categorises risks according to qualitative risk ratings and type of attack being performed. This section also provides a summary of the different ways to batch the simulation of security risks to generate a more comprehensive set of data on which to based security risk decisions.

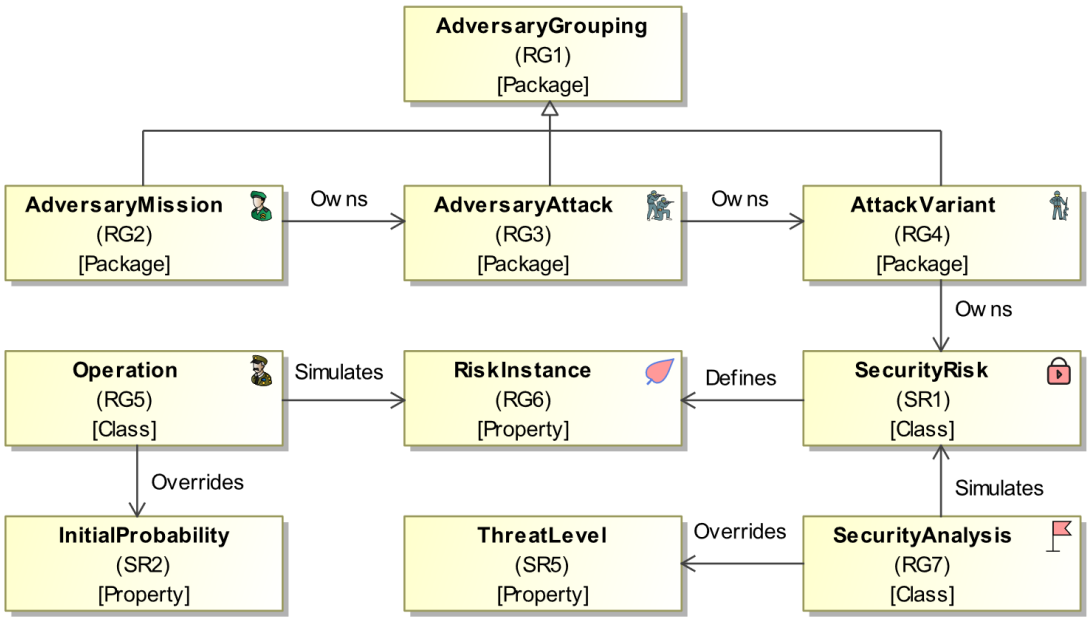


Figure 4-23: *CEMT Security Risk Grouping Taxonomy*

Figure 4-23 shows the taxonomy of stereotypes that the CEMT includes for defining the security risk categories:

SR1. `AdversaryGrouping` – This stereotype inherits from the SysML `Package` metaclass and is a generic stereotype used as a parent for the various objects that categorise the security risks.

SR2. `AdversaryMission` – This stereotype inherits from the `AdversaryGrouping` stereotype and is used as the top-level grouping for security risks. Adversary mission categories are based on the function of the system that the attacker is trying to compromise.

SR3. `AdversaryAttack` – This stereotype inherits from the `AdversaryGrouping` stereotype and is used as the mid-level grouping for security risks. Adversary attacks are subordinate to adversary

missions and are categorised based on the asset that is being compromised by the adversary to deny the function described by the parent mission.

SR4. `AttackVariant` – This stereotype inherits from the `AdversaryGrouping` stereotype and is used as the lowest-level grouping for security risks. Attack variants are subordinate to adversary attacks and represent the different ways that attacker would be able to compromise the asset described in the parent attack.

SR5. `Operation` – The stereotype inherits from the SysML `Block` stereotype and is used to do operation-specific risk simulations. An operation simulates multiple security risks simultaneously, but overrides the `InitialProbability` value within each security risk, so that the modeller can assess the risk in a different threat environment, such as when the system under evaluation is being used for a specific operation or mission.

SR6. `RiskInstance` – The stereotype inherits from the SysML `PartProperty` stereotype and represents an instance of a security risk within an `Operation`.

SR7. `SecurityAnalysis` – The stereotype inherits from the SysML `Block` stereotype and is used to do parallel risk simulations for each possible threat level. A security analysis simulates a single security risk multiple times but overrides the `ThreatLevel` value within each

simulation of the security risk, so that the modeller can assess the risk against all threat actors at the same time.

The adversary missions, adversary attacks and attack variants are used to create a categorisation hierarchy based on the type of attack the adversary is attempting. The missions are the first tier of the hierarchy and represent the goal of the attacker in terms of the system function or capability they are trying to degrade. The attacks are the second tier beneath the missions, and they usually align with a threat against a component or asset within the system that performs a critical part of the function being targeted by the adversary mission. The variants are the bottom tier of the hierarchy, and they represent the different ways in which the attack can be realised. There is a one-to-one relationship between variants and security risks.

This categorisation provides a mission-focused lens through which to view the security risks in the system. This ensures that the information is presented to decision makers in a way that is contextualised to the mission that the system is expected to perform and the ways in which an attack can prevent the system from completing that mission.

Qualitative mappings for likelihood, consequence and resultant risk are also used to categorise risks within the CEMT. Based on the results of the simulation, a security risk is mapped into one of six qualitative likelihood bands – Negligible, Very Low, Low, Medium, High and Extreme. The way in which the quantitative simulation results are mapped into the qualitative likelihood bands is not enforced

by the CEMT and should be based on the organisation's risk governance processes.

The modeller then assigns a qualitative consequence band to the security risk, based on the impact of the risk being realised. These qualitative consequence bands are Minimal, Minor, Moderate, Major, Severe and Catastrophic. Like the likelihood bands, the justification behind allocating these consequence bands is not enforced by the CEMT and is left up to the modeller and their organisation's risk processes. Based on the assigned qualitative likelihood and consequence bands, the CEMT will automatically assign an overall risk rating for the security risk in accordance with Table 4-1, providing a categorisation of all security risks according to which overall risk band they belong.

Table 4-1: *Mapping of Likelihood and Consequence to Risk Rating*

		Consequence					
		Minimal	**Major**	**Moderate**	**Major**	**Severe**	**Catastrophic**
Likelihood	**Extreme**	Moderate	Significant	High	Extreme	Extreme	Extreme
	High	Low	Significant	High	High	Extreme	Extreme
	Moderate	Low	Moderate	Significant	Significant	High	Extreme
	Low	Low	Low	Moderate	Moderate	Significant	High
	Very Low	Low	Low	Low	Moderate	Moderate	Significant
	Negligible	Low	Low	Low	Low	Low	Moderate

To streamline the simulation process, the `SecurityAnalysis` object facilitates the automated analysis of a security risk at all four `ThreatLevels`. The CEMT will automatically create a security analysis for each security risk that is assessed by the tool, and this becomes the primary mechanism by which baseline simulation results are captured within the model. When simulated, the security analysis will run four parallel Monte Carlo analyses on the associated security risk, with the threat level of each parallel simulation forced to be a different threat level. The quantitative results for all four simulations will then be saved onto the security risk object, facilitating the efficient quantitative simulation of the security risks within the model.

Finally, the `Operation` object can be used to re-simulate the security risks for different threat environments. An operation is analogous to a mission or deployment of the system under evaluation and allows a modeller to reassess the risk posed to the system if it is deployed to a location with a higher inherent risk level. The CEMT achieves this by allowing the `Operation` object to instantiate all of the relevant security risks as `RiskInstance` objects and then override the `InitialProbability` of those simulations. This allows the modeller to increase or decrease the likelihood of attacks being attempted between the baseline threat environment used in the standard security analysis and the new threat environment represented by the operation.

This flexibility in running the simulations allows the CEMT risk assessment to have greater relevance through the life of the system, as changes to the design

baseline, threat environment and threat capabilities can quickly be incorporated and re-simulated.

4.5 PROCESS OVERVIEW

The overview of the CEMT provided in this chapter can give the impression that the CEMT process is largely sequential, starting at *Threat Modelling* then moving to *Threat Allocation* and finishing with *Risk Assessment*. However, that sequential structure of those three cardinal phases is primarily used for articulating the mechanics of building the threat model.

In a real-world application of the CEMT, the process will be much more iterative and cyclical. This is because the process is meant to support continuous assessment and assurance of a sustained capability system that is actively maintained throughout the system lifecycle. The following section of the chapter describes a generic process flow for the CEMT which provides insight into the intended implementation of the CEMT within an organisation.

Figure 4-24 depicts a generic process flow for the CEMT. Blue boxes indicate input data, red boxes represent steps the modeller completes within the CEMT process, orange boxes are used to show external libraries, green boxes indicate third party review steps and purple boxes highlight changes within the system being evaluated or its operating context.

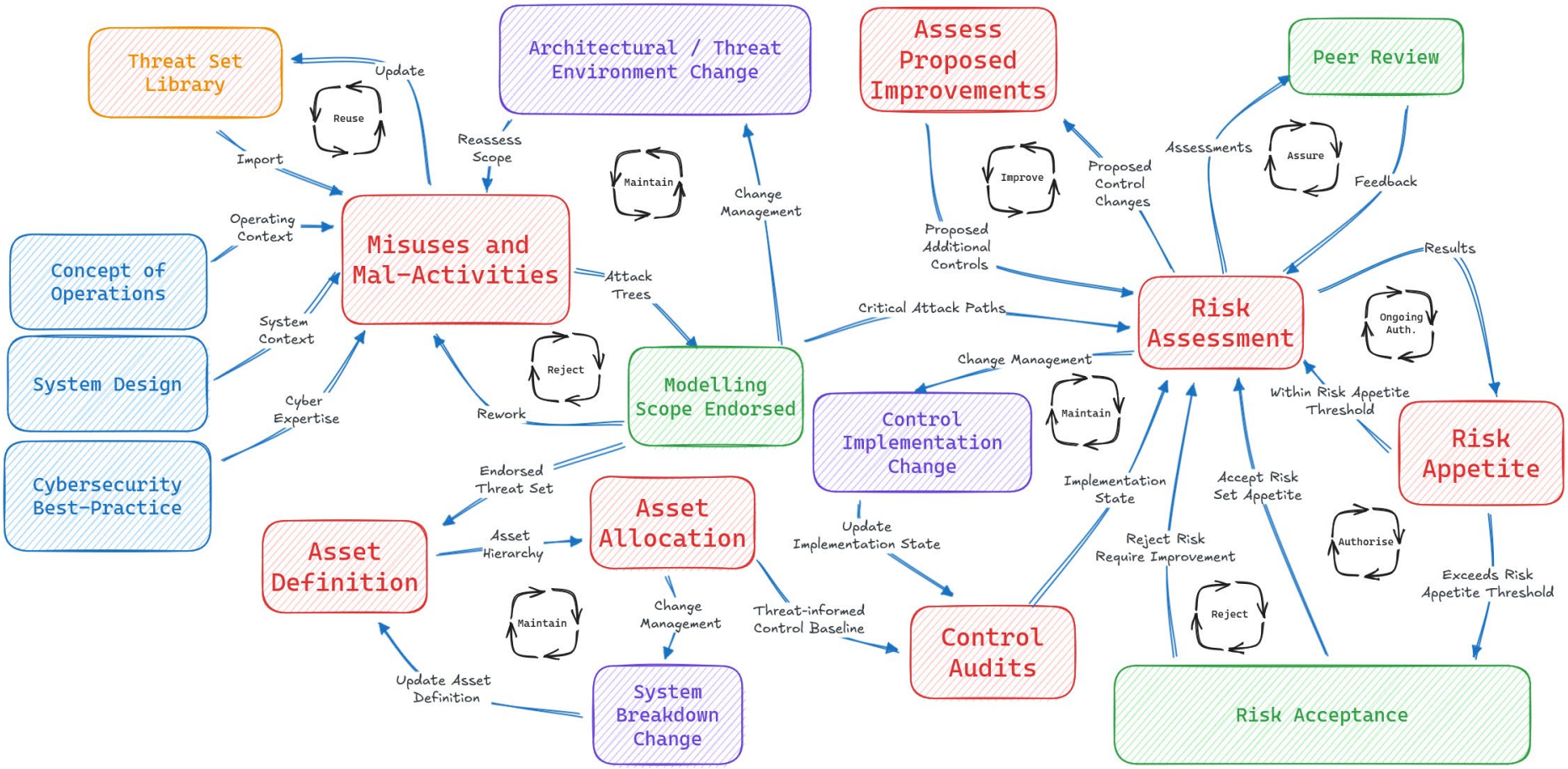


Figure 4-24: *Generic CEMT Process Flow*

The process starts on the left-hand side of the diagram with the three blue boxes which represent the input data to the CEMT process:

- *Concept of Operations* – This is a data set which describes the way in which the system is intended to operate and the critical functions required to perform its mission.
- *System Design* – This is a data set that outlines the design of the system being evaluated. The minimum level of detail required is a context diagram which shows the external interfaces of the system and the ways it interacts with external systems.
- *Cybersecurity Best-Practice* – This is a data set that captures cybersecurity best-practices from reputable sources. This would include cybersecurity control sets such as the Essential Eight [15], ISM 14] and MITRE ATT&CK [121] which can be used to generate the data in the threat model.

These items feed into the development of *Misuses and Mal-Activities*, which broadly aligns with the *Threat Modelling* phase discussed in Section 4.2. Importantly, this step interacts with a *Threat Set Library*, which is a repository of previous threat sets developed within an organisation. The *Misuses and Mal-Activities* step imports existing threat sets from the library and updates the library with any new threat sets that are generated, creating the loop labelled *Reuse* which highlights the reusability which is a key part of the CEMT process.

The *Misuses and Mal-Activities* step produces attack trees which are sent for external review in the *Modelling Scope Endorsed* step. This step focuses on ensuring the scope of the assessment, as defined by the attack surface articulated in the attack trees, is sufficient and appropriate. If it is found to be insufficient, the flow returns via the *Reject* loop for rework. A *Maintain* loop also flows from the *Modelling Scope Endorsed* step, as following this step architectural and threat environment changes must be monitored and any that would impact the endorsement of the attack trees would require a re-assessment of the scope of the CEMT activities.

The *Modelling Scope Endorsed* step provides the endorsed threat set to the *Asset Definition* step, where the asset hierarchy is developed. This asset hierarchy is then linked into the threat model during the *Asset Allocation* step, which generates a threat-informed control baseline. These steps broadly align with the *Threat Allocation* phase discussed in Section 4.3. Maintain loop also loops back into the *Asset Definition* step, as any changes in the system breakdown structure need to be monitored and assessed to determine whether the asset hierarchy needs to be updated over time.

Once the threat-informed control baseline has been derived, the controls within that baseline are audited to determine an implementation state within the system under evaluation in the *Control Audits* step. These implementation states, along with the critical attack paths identified from the attack trees, feed into the *Risk*

Assessment step, which develops and simulates the security risk modules, security risks and security analyses as described in Section 4.4.

The *Risk Assessment* step has its own *Maintain* loop, where changes to the implementation state of controls, due to events like obsolescence replacements, control effectiveness changes and configuration drift, must be monitored and incorporated where appropriate. Additionally, assessments produced in the *Risk Assessment* step flow to the *Peer Review* step, where the detailed assessments are critically reviewed to ensure accuracy and consistency. The feedback loop developed through this peer review process is labelled as *Assure* in Figure 4-24.

Another important step that flows out of the *Risk Assessment* step is the *Assess Proposed Improvements* step, where the modeller uses the Proposed implementation state on the controls to recommend additional controls that could reduce the residual risk in the system. These get simulated in parallel to the baseline assessment to produce a proposed/residual risk that can be compared to the inherent/current risk that is determined in the baseline assessment. The *Assess Proposed Improvements* step produces a set of proposed additional controls, and this completes the *Improve* loop shown in Figure 4-24.

Ultimately, the results of the *Risk Assessment* step are passed to the *Risk Appetite* step, where the resultant risks are compared against the approved appetite thresholds defined within the model. Where those appetites have already been approved and the assessed risk is within that appetite, the flow loops back to the *Risk Assessment* step, representing continuous assurance within

acceptable limits. This is labelled as the *Ongoing Authorisation* loop. When the assessed risk exceeds those appetite thresholds, or when those appetite thresholds have not been approved because it is the first time this system has been assessed, the flow progresses to the *Risk Acceptance* step.

The *Risk Acceptance* step involves a presentation of the risk results to organisational decision makers who either accept the residual risk and set risk appetite thresholds for each risk, following the *Authorise* loop, or they reject the risk and require additional mitigations to be implemented, following the *Reject* loop.

The result of this generic CEMT process flow is a sustainable and maintainable threat model that produces assured risk assessments the minimise bureaucratic overhead while ensuring that risks are accepted by the appropriate authority, when required. While the specific CEMT process flow within an organisation is likely to look different to that shown in Figure 4-24, because most organisations have slightly different risk governance approaches, the intent of this generic flow is to show that a proper implementation of the CEMT should be iterative and cyclical.

CHAPTER 5 FACE VALIDITY TRIALS

5.1 EMBEDDED CHAPTER – MDPI SYSTEMS

This chapter is published as

Fowler, S.; Joiner, K.; Ma, S. Cyber Evaluation and Management Toolkit (CEMT): Face Validity of Model-Based Cybersecurity Decision Making. *Systems* **2024**, *12*, 238. https://doi.org/10.3390/systems12070238

systems

MDPI

Article

Cyber Evaluation and Management Toolkit (CEMT): Face Validity of Model-Based Cybersecurity Decision Making

Stuart Fowler [1,*], Keith Joiner [2] and Siqi Ma [1]

[1] School of Systems and Computing, University of New South Wales, Canberra 2600, Australia; siqi.ma@unsw.edu.au
[2] Capability Systems Centre, University of New South Wales, Canberra 2600, Australia; k.joiner@unsw.edu.au
* Correspondence: stuart.fowler@unsw.edu.au

Abstract: The Cyber Evaluation and Management Toolkit (CEMT) is an open-source university research-based plugin for commercial digital model-based systems engineering tools that streamlines conducting cybersecurity risk evaluations for complex cyber-physical systems. The authors developed this research tool to assist the Australian Defence Force (ADF) with the cybersecurity evaluation of complicated systems operating in an increasingly contested and complex cyber environment. This paper briefly outlines the functionality of the CEMT including the inputs, methodology, and outputs required to apply the toolkit using a sample model of the process applied to a generic insider threat attack. A face validity trial was conducted on the CEMT, surveying subject-matter experts in the field of complex cybersecurity analysis and risk assessment to present the generic case study and gather data on the expected benefits of a real-world implementation of the process. The results of the face validity broadly supports the effectiveness and usability of the CEMT, providing justification for industry research trials of the CEMT.

Keywords: cybersecurity; model-based systems engineering; CEMT; cyber; digital engineering; mission engineer; cyber-worthiness; model-based risk assessments; threat modelling

Citation: Fowler, S.; Joiner, K.; Ma, S. Cyber Evaluation and Management Toolkit (CEMT): Face Validity of Model-Based Cybersecurity Decision Making. *Systems* **2024**, *12*, 238. https://doi.org/10.3390/systems12070238

Academic Editors: Gregory S. Parnell and Eric Specking

Received: 27 May 2024
Revised: 18 June 2024
Accepted: 29 June 2024
Published: 1 July 2024

1. Introduction

1.1. Background

Making informed and accurate decisions on the acceptability of cybersecurity risk to complicated engineered systems is a challenging problem. There is an understandable urge to try and simplify the problem to a purely technical challenge by mandating a set of technical-, procedural-, and personnel-focused security controls and auditing the compliance to that baseline set of security controls. However, this reductionist approach often neglects the complexity of the system's operating context, mission, and safety criticality, especially where organisational boundaries separate aspects like estate infrastructure from ICT infrastructure and user platforms (i.e., ships, aircraft, vehicles, etc.). Not all controls are relevant to all systems, and not all systems require the same level of assurance when it comes to mitigating cyber threats. Contemporary security assessment methodologies [1,2] have introduced risk-based decision-making processes to allow compliance frameworks to be applied. The specific steps vary between the various methodologies, but the generic approach is that the results of a compliance audit and control effectiveness determination are used as an input to a qualitative risk decision made by an appropriately authorised decision maker.

Figure 1 depicts a generic process for how these risk-based decisions are made. At step 1 in the top right corner, a system is designed with security being one of the quality factors driving that design. A security assessment is then made, typically by auditing a set of best practice mitigation controls (step 2) such as the ISM [3], NIST 800-53 [4], or similar, and assessing the residual risk based on the level of compliance found during that audit

(step 3). This residual risk determination includes an implicit and qualitative analysis by the security assessor of the effectiveness of any compliant controls and the importance of any non-compliant controls. The residual risk is then provided to a decision maker, who holds a level of authority within the organisation that allows them to accept the risk, and they compare this against the organisation's risk tolerance. Based on this comparison, they either accept the risk, authorising the system for operation, or reject the risk, requiring further mitigation of the residual risk by the design team before the system can be accepted into operation (step 4).

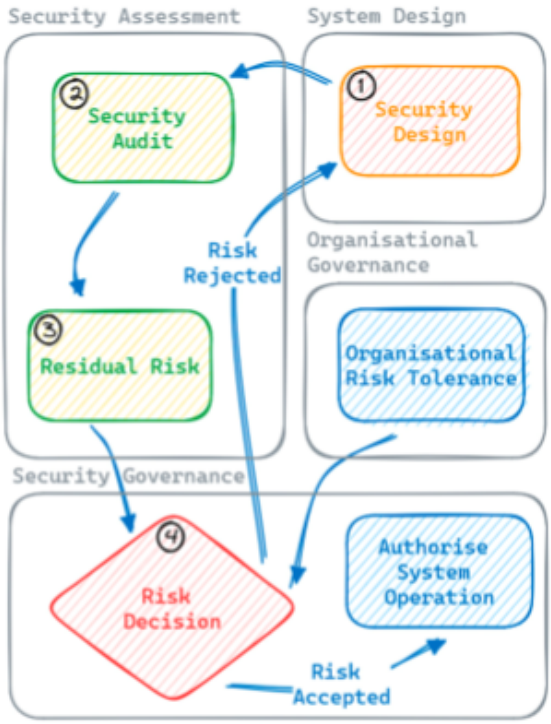


Figure 1. Generic process for making cybersecurity risk decisions under a compliance-driven risk assessment paradigm.

One of the challenges introduced by this approach is that it is reliant on the control baselines used for the audit being directly relevant to the system being audited. The security assessment usually includes a step to exclude controls that are not applicable, but any non-compliances against applicable controls are often difficult to assess objectively because the control baselines are often not prioritised within the context of the system being assessed. The individual security assessor then needs to implicitly assess the importance of a particular control non-compliance and incorporate that into the residual risk. This process is often subjective and opaque, which can make it difficult to explain and justify to other participants in the risk assessment that do not have cyber expertise. Consequently, the decision maker usually either needs greater cyber expertise to understand the assessment, or they need to blindly trust the assessor.

Compliance-driven approaches can leave decisions makers in a position where it is difficult to understand the level of risk being accepted, and consequently hamper effective risk decisions. The need for a better way to describe and assess risk has been identified in contemporary military doctrine, as evidenced by the following statements from the U.S. Department of Navy (DON):

> *"The DON needs a new approach to cybersecurity that goes beyond compliance because our over-reliance on compliance has resulted in insecure systems, which jeopardise the missions these systems support. Instead of a compliance mindset, the DON will shift to*

> *Cyber Ready, where the right to operate is earned and managed every day. The DON will make this transition by adhering to an operationally relevant, threat-informed process that affordably reduces risk and produces capabilities that remain secure after they have been delivered at speed".* [5] (p. 7)

The reliance on the opaque expert judgement of the assessor is exacerbated because these assessments are typically documented in spreadsheets [6,7] and databases [8], which do not effectively describe the system's security risk to a layman. Moreover, such spreadsheets can be difficult to maintain and update as new threats emerge throughout the capability sustainment life cycle, as the underlying control baseline is updated over time and as the capability design evolves through development and in-service. Emerging techniques [8] use 'compliance-as-code' techniques to improve reusability and to reduce the difficulties of amalgamating assessments at a system-of-systems level. However, these techniques primarily address the amalgamation of a compliance level, rather than the amalgamation of residual risk levels, meaning that the challenges of opaque expert judgement remain.

A 'best-practice control' set is more easily achieved in a predominantly information communications technology (ICT) system, as there is a uniformity in ICT architectures compared to cyber-physical systems, or what ICT professionals commonly refer to as the 'edge'. However, the 'edge' forms an integral part of the cyber-attack surface and the kill-chain effects of complex engineered systems, which means that the direct application of the best-practice control sets derived from that comparatively uniform ICT environment can lead to a focus on the wrong type of controls. Additionally, the expert judgement of importance and priority performed by assessors can be influenced by an implicit bias towards those uniform ICT architectures. The best-practice control sets help immensely in terms of assessment efficiency within that uniform environment; however, the cyber-physical environment often requires a return to first-principles security and engineering to obtain an accurate cybersecurity assessment. First-principle cybersecurity assessments of engineered systems are especially useful in developing risk mitigations to recommend.

1.2. Literature Review

The existing alternatives in the literature to this compliance-driven risk assessment are threat-based risk assessments. Threat-based risk assessment methods provide better flexibility in terms of applying the process to different system contexts at the expense of requiring a higher level of expertise and understanding from the assessors.

The widely known 'MITRE ATT&CK' framework [9] facilitates threat-based assessment, but it is more of a taxonomy and repository of known behaviours to understand threats and attack patterns, rather than an evaluation and assessment methodology. While the framework derives mitigations and detections for implementation in systems, a direct application of these would essentially be another compliance-driven assessment, which can suffer from the same challenges introduced earlier.

Similarly, Mission-based Cyber Risk Assessments (MBCRA) such as Cyber Table Top [10] and Mission-based Risk Assessment Process for Cyber (MRAP-C) [11], required under the U.S. Department of Defence acquisition processes, provide an example of threat-based approaches to deliver risk assessments. However, an analysis of the MBCRA implementations concluded in 2022 that the explainability of the product of the MBCRA can still be limited due to the data output format [12]. The MBCRA process is inherently collaborative, which improves the situation in terms of opaque expert judgement, whereby decision makers can rely on the tempering and recording effects of group consensus. However, the limited explainability still constrains the goal of truly informed cyber risk decision making.

The literature also outlines approaches that may help with the explainability of cybersecurity assessments. According to Kordy et al. [13], *"the great advantage of graph-based approaches lies in combining user friendly, intuitive, visual features with formal semantics and algorithms that allow for qualitative and quantitative analysis"*. The intuitive nature of these approaches shows promise for facilitating collaboration with stakeholders who may not have a deep cybersecurity background. Kordy et al. [13] categorised graph-based security

analyses into two broad categories: statically modelled graphs and sequentially modelled graphs. Both categories can deal with either just attacks or both attack and defence.

Attack trees, first outlined in 1999 by Weiss [14] and popularised by Schneier [15], are the basis for many of the statically modelled graphs used for cybersecurity analysis. These graphs depict a hierarchical tree structure to describe the combination of events that need to occur for a cybersecurity risk to be realised and are broadly similar to the fault trees common to system safety analysis. Attacks trees have been successfully incorporated into the likelihood calculations of traditional risk assessment methodologies from 2014 [16], and have been extended from 2016 to include the depiction of defensive actions [17] and termed attack-defence trees. In 2015, the risk likelihood assessments that flowed from attack trees were estimated by converting the attack tree structure into a Bayesian network in order to combine probabilities of the various nodes on the attack tree [18,19].

The sequentially modelled graphs outlined in the literature are comprised of a sequence of events that an attacker must perform in order to compromise a target system. This approach provides a focus on the intent of an attacker; one of the attributes missing in the standard systems security approaches and a fundamental advantage when the threat is evolving quicker than the system. Inherently, such approaches can focus the assessment on innovative mitigations to limit the likelihood of successful attack. The concept of insecurity flow graphs was introduced early by Moskowitz and Kang [20] to model this attack sequence, though this concept attempted to reduce the graph into a mathematical model, losing some of the intuitiveness benefits of a graph-based approach.

The construction of misuse case graphs, also known as abuse case graphs, is another approach in the literature that focuses on the actions of the attacker and their interaction with the target system. Appearing in the literature at the turn of the century in McDermott [21] and Sindre et al. [22], misuse cases borrow from the use case concept common in software development and systems engineering practices but focus on how an adversary, rather than an authorised user or operator, might interact with a system. More recent extensions to the misuse case concept include overlaying the misuse cases with the system architecture [23] and developing misuse case coverage metrics as a proxy for security risk [24]. On their own, misuse cases require extensive textual encoding to adequately convey the full risk picture [25].

While misuse cases are often used to elicit security requirements [26], mirroring the purpose of use cases in software development, Matuleviaius [27], in 2017, linked misuse cases to the mal-activity graphs proposed by Sindre [28]. These mal-activity graphs provide a sequential model of the activities taken by the attacker—much like the insecurity flow graphs introduced by Moskowitz and Kang [20]—and can be used to provide additional detail to support misuse case graphs.

The literature does include some comparison studies between the attack tree methodologies and the misuse case methodologies. Both Opdahl et al. [29] and Karpati et al. [30] found that while attack tree methodologies tended to identify more threats, there were significant differences between the types of threats found by each technique. This would suggest that a combined approach such as that outlined by Tondel et al. [31], which looks at both misuse cases and attack trees, might be the most effective approach.

Efficiency in risk assessment is also an important concern, particularly for organisations with many legacy cyber-physical systems to assess that have only recently exploited ICT networking (aka Internet of Things (IoT)). While there has been some literature looking to increase the reusability of these graph-based threat models [32] these techniques still require knowledgeable experts to apply them to systems individually. However, in the development of bespoke, complex cyber-physical systems, the technical knowledge of those involved with the assessment is not necessarily the key limitation of a successful risk assessment—the ability to explain that technical risk assessment to a decision maker who understands that the broader operational and strategic context can be the issue.

1.3. Cyberworthiness

'Cyberworthiness' is a term that was created recently within the Australian Department of Defence and invokes existing worthiness concepts under which Defence is responsible for internally regulating the safe use of their systems. The existing worthiness systems began in the 1990s with airworthiness regulations and governance [33,34], which led to seaworthiness [35] and landworthiness [36] being defined and enforced. 'Worthiness' requires a level of regulation and monitored assurance over design, deployment, and sustainment. As 'worthiness' is often synonymous with continued safety and functionality, these regulations are applied more rigorously based on safety criticality.

Cybersecurity has challenged Australian Defence, as it has the U.S. Defence, though Australia is arguably lagging in this realm [37–40]. As cybersecurity threats became more prevalent and necessary across more safety-critical systems that do not fit into the traditional definition of an ICT system, it became necessary to apply broader and more enforceable assurances than those applied to generic ICT systems. Consequently, cyber-worthiness was first termed [41,42] and is defined as:

> *"The desired outcome of a range of policy and assurance activities that allow the operation of Defence platforms, systems and networks in a contested cyber environment. It is a pragmatic, outcome-focused approach designed to ensure all Defence capabilities are fit-for-purpose against cyber threats".* [43]

The inclusion of the worthiness construct is a deliberate reference to the established technical assurance frameworks that underpin concepts such as airworthiness and seaworthiness. This requires a level of rigour in the engineering analysis that provides confidence and assurance that the claim of worthiness is accurate, complete, and supported by evidence. However, a worthiness claim must also be understandable by those making decisions about the operation of a system, as outlined by the Principles for Governance of Seaworthiness from the Defence Seaworthiness Management System Manual [44]:

> *"2.10 The seaworthiness governance principles require that seaworthiness decisions are made:*
>
> *a. mindfully—decisions are more effective and less likely to have unintended consequences when they are made with a thorough understanding of the context, the required outcome, the options available, and their implications now and in the future*
>
> *b. collaboratively—obtaining input from all stakeholders and engaging in joint problem-solving results in better decisions (bearing in mind that collaboration does not necessarily require consensus)*
>
> *c. accountably—decisions only become effective when people take accountability for making them happen*
>
> *d. transparently—decisions are more effective when everyone understands what has been decided and why".* [44] (p.33)

Cyberworthiness then requires not only the rigorous application of cybersecurity analysis, but also demands the explicit inclusion of a broad range of stakeholders and a focus on genuinely informed decision makers to ensure that they can reasonably take accountability for the decisions they are asked to make.

1.4. Addressing the Problem

The solution to the identified challenges is not as simple as just implementing academic approaches. These approaches are not new, with some almost 25 years old, but they do not have the industry traction in engineered systems development. The primary concerns that limit the direct application of these techniques to complicated cyber-physical systems with evolving complex threats are:

- Usability—there is limited ability to easily create and review these graph-based threat assessments, especially in large, complex systems;

- Efficiency—reusability of these assessments is limited in comparison to compliance-based approaches that re-apply a common control set;
- Maintainability—it is difficult to update complex graph-based assessments without specialised toolsets as the system or threat environment evolves.

In particular, sequential graph methods can be difficult to understand at a macro level, while static graph methods can be unwieldy and difficult to maintain as the system changes. However, structured languages such as those in model-based systems engineering (MBSE) may be able to be used to unify sequential and static graph methods so that the security designers and evaluators can produce one set of graphs and automatically generate views of other graphs based on the relationships defined in the modelling process. The inherent strengths of model-based approaches can also help maintainability through the object-oriented nature of the dataset, allowing for rapid update and consistency checks of complex views.

It is our thesis that a rigorous and streamlined threat-based approach to security analysis that describes threat vectors and risks using graphs and leverages model-based systems engineering techniques to support maintainability can deliver a methodology that overcomes many of the cybersecurity assessment challenges. Our thesis can be evaluated by delivering such a methodology to project teams who face these challenges when using the traditional compliance-based risk assessments currently implemented in government and industry. This threat-based approach would align with the overall concept of a mission-based cyber risk assessment but drive the implementation of that concept towards one that delivers an intuitive and maintainable output of threat graphs for through-life cyber resilience. Such an approach likely helps cybersecurity assessors by leveraging the systems and safety engineering culture and competencies in government and industry.

This threat-based approach would also address the decision-making principles underpinning cyber-worthiness. The approach must be mindful of the context in which the system is operating, be intuitive enough to facilitate collaboration between all stakeholders, be understandable by key decisions makers so that they can reasonably take accountability for their decisions and be transparent in terms of how the conclusions trace to the analysis.

Our research aimed to determine whether this more extensive threat-based methodology and associated toolset would help improve cybersecurity risk assessments across more diverse systems, especially cyber-physical systems at the 'edge', in comparison to traditional compliance-driven approaches. Such improvements would be categorised across dimensions of effectiveness, efficiency, and usability to understand both the benefits and costs associated with an alternative approach. Seeking feedback on the methodology and associated toolkit through industry workshops is key to ensuring that practitioners can provide a preliminary vetting of any proposals before allocating resources to costly implementation trials. Structured workshops with representative domain experts to examine the early and likely validity of the method can be referred to as 'face validity' and has been debated academically by Allen et al. [45].

Our primary research question asked:

- Are integrated threat models, developed using model-based systems engineering (MBSE) techniques, an effective and efficient basis for the assessment and evaluation of cyberworthiness?

To help answer this primary question, we derived the following three sub-questions that characterised effective cyberworthiness based on how informed the decision makers were, efficiency in terms of the reusability of the assessment effort, and maintainability of the assessment through-life as well as a qualitative question seeking preference data from practitioners:

- Do the developed threat models provide decision makers with the necessary understanding to make informed security risk decisions?
- Does the process provide sufficient reusability and maintainability that the methodology is more efficient than prevailing compliance-based approaches?

- Do cybersecurity risk practitioners prefer the integrated threat model approach to traditional security risk assessment processes?

2. Materials and Methods

Our methodology for investigating this problem set requires a level of development of a proposed security assessment approach that brings together threat-based analysis, graph-based presentation techniques, and model-based implementations.

This methodology began by determining what we needed our threat-based analysis to take as inputs and produce as outputs, which we defined as threat-based cybersecurity engineering. We then began prototyping implementations of this approach using model-based systems engineering techniques to iteratively identify the features that worked and any challenges that needed to be overcome [46]. We refined this prototype to develop a candidate toolset for implementation and evaluation [47]. This resulted in the creation of the Cyber Evaluation and Management Toolkit (CEMT), which included a generic example implementation of a common threat to complex systems [48]. This example implementation was then presented at conference tutorials and industry briefings with feedback from participants received via survey [49]. This provided a result set that described the face validity of the approach, outlining preliminary impressions from industry and academic experts as to the feasibility and expected benefits of a wider implementation of the approach.

This section provides a summary of the work carried out to define the threat-based cybersecurity engineering concept and develop the Cyber Evaluation and Management Toolkit (CEMT) alongside a high-level summary of the generic example implementation used for the face validity surveys. The results of these face validity surveys are then presented in the results section of this paper, with a discussion of the results at the end of this paper, which will also outline future work to conduct implementation trials on complicated engineered systems in the real world to finalise the evaluation of the developed approach.

2.1. Threat-Based Cybersecurity Engineering

In order to guide the development of the CEMT, it was necessary to first articulate an evaluation framework in which the toolkit would be operated. For this reason, our threat-based cybersecurity engineering concept was created, and the associated inputs and outputs were articulated.

Threat-based cybersecurity engineering is defined as the process by which a system's capability is continuously assessed to ensure that relevant cybersecurity mechanisms are incorporated into the capability design and the resultant residual cybersecurity risks are accurately determined and explained. This definition was developed within the context of cyberworthiness and the need for an approach to cyber evaluation that provides rigorous assurance to underpin and justify any cybersecurity recommendations.

Threat-based cybersecurity engineering is primarily aimed at securing complicated cyber-physical systems that require a first-principles approach to identifying how to appropriately secure a system from complex threats. It is threat-based and context-focused and derives a set of relevant security controls that should be incorporated within a system. It builds on the systems security engineering processes defined in [50] and incorporates the framing and categorisation of the NIST Cybersecurity Framework [51]. It is a collaborative process that brings together experts from different parts of an organisation and facilitates meaningful discussion through intuitive analysis and quantitative reasoning. The cybersecurity engineering process is one of facilitation, which seeks to incorporate the expertise of various stakeholders and present assessments in a transparent, traceable, and explainable manner, rather than being reliant on cybersecurity polymaths to conduct opaque and ultimately subjective assessments.

The generalised inputs and outputs to and from threat-based cybersecurity engineering are depicted in Figure 2.

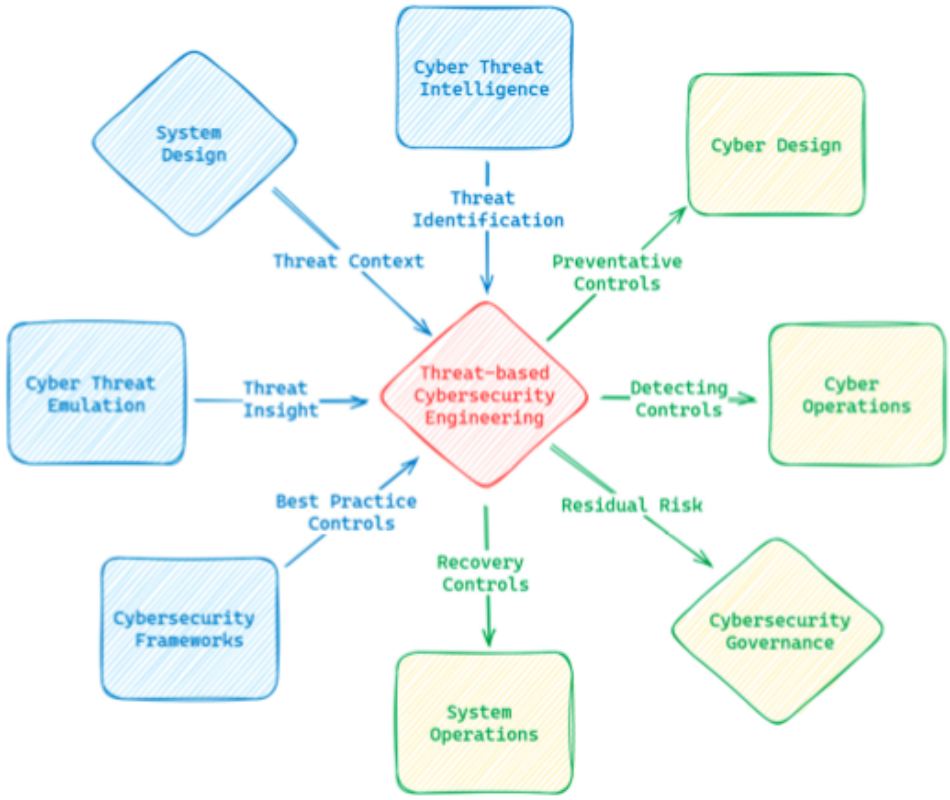


Figure 2. Threat-based cybersecurity engineering context.

The primary inputs to threat-based cybersecurity engineering are:

- Threat Context, derived from the system or capability design/architecture;
- Threat Identification, provided by the Cyber Threat Intelligence function within an organisation;
- Threat Insight, contributed by the Cyber Threat Emulation function within an organisation;
- Best Practice Controls, distilled from the various cybersecurity frameworks available within the cybersecurity body of knowledge.

The Threat Context input represents the functions, interfaces, and components within the system or capability that is being evaluated. This defines the threat surface for the cyber evaluation and informs both the scope and structure of the threat model. The design of the system and the missions which the system performs are the primary source of the threat context.

The Threat Identification input is the attack surface of the system, essentially the set of attack vectors and the threat actors that perform them, which are relevant to the system being evaluated. This is typically contributed by the Cyber Threat Intelligence function, which monitors and identifies potential threat actors, and can consist of both generic threat sets that cover common threats to information systems such as network penetration threats and supply chain attacks, and more specific advanced threat sets such as electronic attack or signal processing exploitation.

The Threat Insight input describes the tactics, techniques, and procedures that would be used by the identified threat actors to achieve their malicious intent. This encompasses the threat vectors by which the system might be attacked and articulates the steps that must be taken by a threat actor in order to compromise the system being evaluated. The Cyber Threat Emulation function, which maintains a detailed understanding of the methods used by threat actors to access systems, is the key contributor of the Threat Insight input.

The Best Practice Controls input encompasses the corpus of information that captures the cybersecurity mitigation mechanisms that can be used to prevent, detect, respond to and recover from a cyber-attack. Importantly, the intent here is to not limit the evaluation to the application of a single control framework audited to determine compliance, rather, the superset of control frameworks is distilled by the cybersecurity engineer when developing and reviewing the threat model, and controls are selected based on their applicability to that threat model, rather than their providence within a particular control framework. A rigorous methodology must simultaneously consider control sets from NIST [2,38], ASD [3], MITRE [9], and others, and select them based on relevance, rather than aligning to any one specific framework.

The primary outputs from threat-based cybersecurity engineering are:

- Preventative Controls, a baseline of preventative cybersecurity controls within the system, for inclusion in the system design;
- Detecting Controls, a baseline of detection and response controls relevant to the system, for implementation by the Cyber Operations function within an organisation;
- Recovery Controls, a baseline of recovery and resilience controls relevant to the system, for implementation by the System Operations function within an organisation;
- Residual Risk, the overall risk presented by the threats to the capability given the mitigation mechanisms that are in place.

The Preventative Controls output encapsulates the mitigation mechanisms that can be applied to the system components to prevent a threat vector from being successfully completed. These are implemented into the cyber design of the system or capability and the level of implementation of these controls drives the overall likelihood of a cybersecurity risk being realised.

The Detecting Controls output includes the mitigation mechanisms that can be incorporated into the system to identify an attack on the system and respond to that attack. These controls are allocated to the Cyber Operations function of an organisation, where analysts are responsible for security monitoring and incident response for the capability being evaluated.

The Recovery Controls output describes the mitigation mechanisms that can contribute to the ability to recover from a successful cyber-attack. These controls are incorporated by the System Operations function of an organisation, where personnel are responsible for ensuring that the system or capability is operational and performant. The implementation of these controls can mitigate the consequence of a cybersecurity risk being realised.

The Residual Risk output represents the overall risk that exists within a system or capability, based on the threats to the system and the control mechanisms in place to mitigate that threat. This residual risk is passed to the Cybersecurity Governance function within an organisation, where it is compared against the organisations' risk appetite to determine whether the level of residual risk is acceptable.

The threat-based cybersecurity engineering artefacts should be expressed using threat graphs that provide explicit traceability between the inputs and outputs. This facilitates the justification of the outputs, based on the inputs, to provide assurance of the relevance of the control set that is ultimately derived. It must also be documented in a way that is agile and responsive to change so that the assessment artefacts can be iteratively updated throughout the system life cycle, ensuring that the outputs of the process remain evergreen and the overall risk management outcomes are enduring.

2.2. Cyber Evaluation and Management Toolkit (CEMT)

Our research then involved the development of a methodology that satisfied this threat-based cybersecurity engineering approach using model-based systems engineering as a potential candidate for solving the challenges identified earlier in this paper. Our early prototype work [46] set the scene for the adaptation of SysML diagrams such as use case diagrams and activity diagrams to model misuse cases and mal-activity flows in the target system. The threat models developed using these views were then used as a lattice for

tracing security mitigations and a map for assessing the priority risks in the system. This high level approach is depicted in Figure 3.

Figure 3. High level approach for our model-based implementation of threat-based cybersecurity engineering.

The initial, informal feedback we received on the prototype was that it required too much focus on how to build the models, rather than the outputs of the assessment. These usability concerns were the primary driver of a refinement activity where a custom plugin was developed to abstract away the complexities of the underlying model-based systems engineering tools and of SysML as a language. The result was the Cyber Evaluation and Management Toolkit (CEMT), which is a plugin to the CATIA Magic software package that implements a taxonomy of custom stereotypes and a suite of custom diagrams to streamline the process initially developed in our prototype work [47].

The CEMT is freely available, and the plugin, source code, and detailed documentation can be publicly accessed on GitHub [48]. The CEMT development included the creation of a generic worked example of two misuse cases, Physical Incursion and Insider Threat, which is also distributed with the toolkit as a sample implementation. This worked example was presented at academic conference tutorials and professional industry workshops, and survey responses were sought from the participants to assess the face validity.

This worked example is summarised in the following section to provide context to the results presented later in this paper.

2.3. CEMT Sample Model

This section outlines the worked example that is provided as a sample implementation with the CEMT. This sample model provides a generic assessment of the Insider Threat and Physical Incursion misuse cases of a generic system. It is not meant to be an exemplar implementation, but is rather meant to illustrate the overall methodology, provide an instructive example of how to develop models using the CEMT, and demonstrate the outputs that are created by the plugin.

The summaries in this section are not meant to be comprehensive descriptions of an entire threat vector and are deliberately kept high-level for brevity, and example diagrams were chosen for readability in this format. All diagram types, along with a comprehensive exploration of the low-level process for creating these diagrams, are available for viewing by the reader in the online documentation available in GitHub [48], where complex diagrams can be more easily displayed and explored.

2.3.1. Threat Modelling

The Threat Modelling phase is the starting point of the CEMT modelling approach and comprises the following steps:

1. Misuse case diagrams;
2. Intermediate mal-activity diagrams;
3. Detailed mal-activity diagrams.

The first step in this phase involves the creation of a CEMT misuse case diagram, which is a customised version of the SysML use case diagram. This diagram is used to define the top-level threats to the system under evaluation by using misuse cases to represent those top-level threats. Threat actors that perform the misuse cases are also placed on these diagrams and linked to the misuse cases they perform. An example of a misuse case diagram is shown in Figure 4.

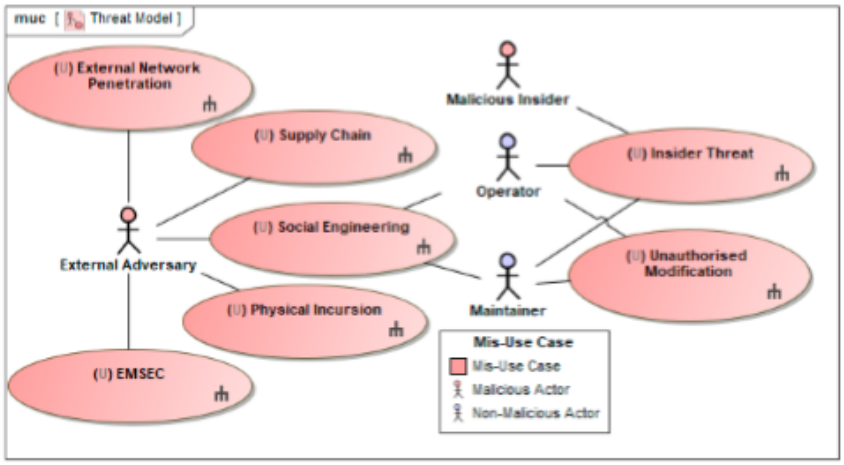


Figure 4. CEMT misuse case diagram from the generic sample model.

For each misuse case, a CEMT mal-activity diagram is then created, which articulates the various threat vectors that could be used to achieve the top-level threat. Figure 5 depicts an example of this for the Insider Threat misuse case, where the steps required to perform this attack are defined. These steps can each have another CEMT mal-activity diagram below them to describe the more specific details. This results in a nested series of flowcharts that show the actions the attacker needs to take as well as the ways in which a defender would be able to detect the attackers' activity. This nested behaviour can be seen in Figure 6, which depicts the internal detail of the Exploit System Access block from Figure 5 with the input and output pins on each figure representing the same object, allowing for the threat flow to pass between diagrams at different levels of the nested hierarchy.

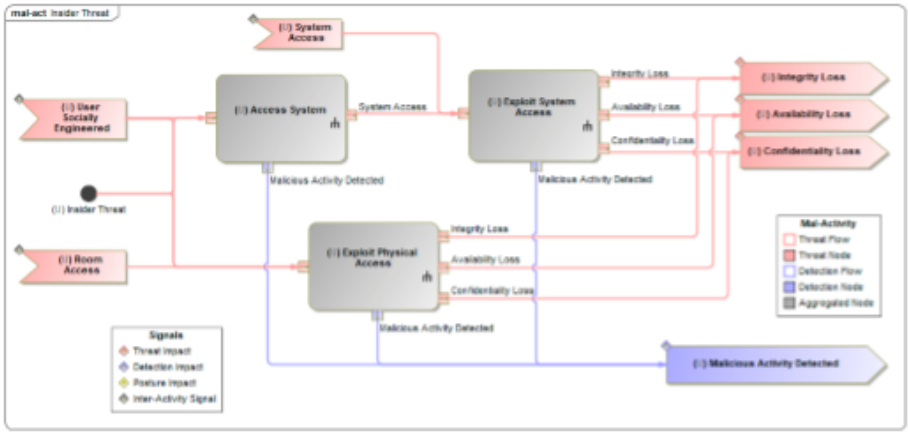


Figure 5. Intermediate CEMT mal-activity diagram for the Insider Threat misuse case.

These nested mal-activities diagrams continue until they reach the bottom level of the hierarchy necessary to articulate the elemental steps that the attacker needs to take to complete the attacks. An example of the lowest level detailed mal-activity diagram is shown in Figure 7, which depicts the internal detail of the Access Sensitive Data block from Figure 6. These detailed diagrams also outline the ability of these threat actions to be detected.

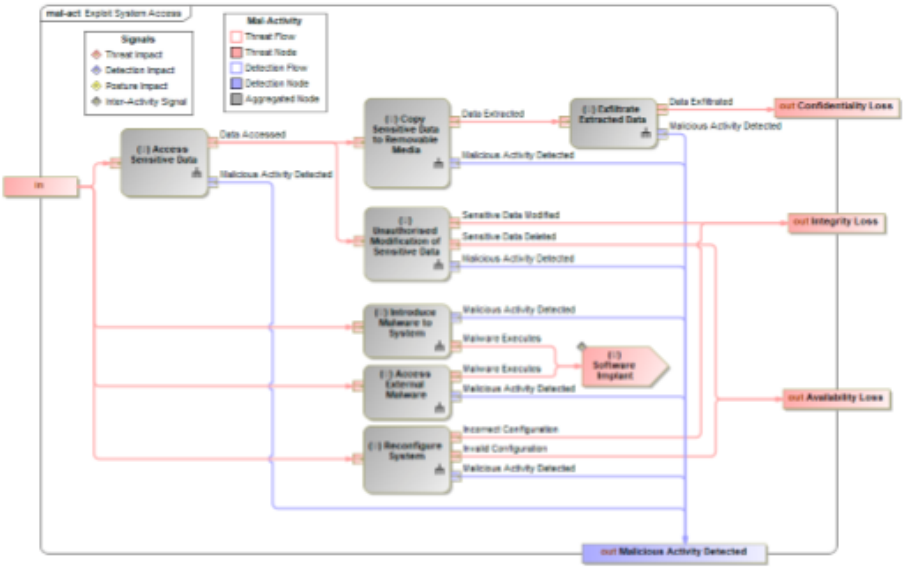


Figure 6. Intermediate CEMT mal-activity diagram for the Exploit System Access mal-activity.

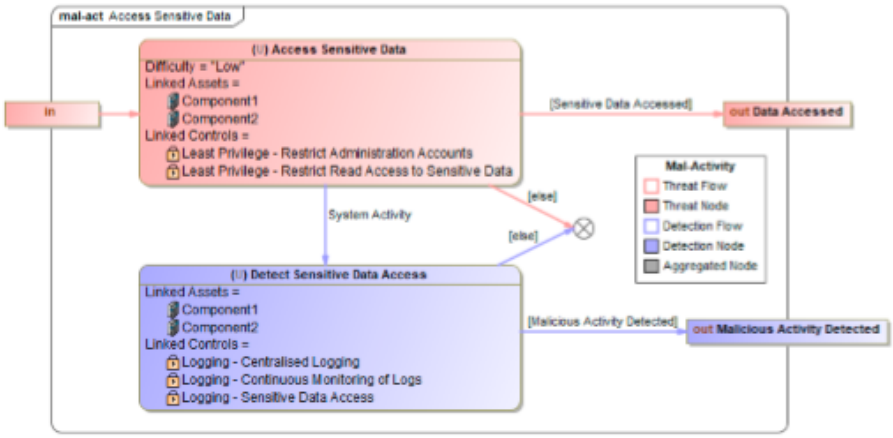


Figure 7. Detailed CEMT mal-activity diagram for the Access Sensitive Data mal-activity.

Following the completion of this Threat Modelling phase, a comprehensive threat model that identifies the threats to the system and the threat vectors through which those threats could compromise the system was developed.

2.3.2. Threat Mitigation

The next phase in the CEMT modelling approach is the Threat Mitigation phase, which comprises the following steps:

1. Allocating assets to the threat model;
2. Tracing controls to the threat model.

The first step involves reviewing the threat nodes and detection nodes in the detailed CEMT mal-activity diagrams in the threat model and identifying the components in the system that are affected by those actions. These components are then created as Asset objects and linked to the specific steps of the threat model. This is shown in Figure 7, where the Component1 and Component2 assets were linked to the relevant nodes of the threat model as they are the components of the system that contain the sensitive data that could be accessed in these actions.

The next step involves reviewing the same actions in the threat model and proposing controls that could potentially be implemented to prevent the threat action from being completed. These are created as objects and linked into the threat model. These controls should be sourced from best-practice frameworks and the knowledge and experience of the modeller. Importantly, the intention is not only to identify controls that are implemented in the system, or even just controls that should be implemented in the system. Ideally, any control that could feasibly be implemented in the system and would directly mitigate the likelihood of the action being completed by the threat actor should be created and linked to the specific step. Figure 7 shows both the preventative controls linked to the relevant threat node and detecting controls linked to the relevant detection node.

The CEMT then instantiates each of the linked security controls onto the linked assets. This creates a list of asset-allocated mitigation mechanisms that are derived from the threat model, ensuring that these controls to be audited are both relevant to the system and are directly contributing to the mitigation of a threat vector. The implementation state of these instantiated objects can be audited and then input into the model to complete the Threat Mitigation phase.

2.3.3. Risk Assessment

The final phase in the CEMT modelling approach is the Risk Assessment phase, which comprises the following steps:

1. Attack tree assessment;
2. Parametric risk analysis;
3. Risk evaluation.

This phase involves the modeller working with various system stakeholders to review the threat vectors, identify those that present the highest risk, and then construct simulation analyses of those high-risk threat vectors to produce quantitative data to support an overall residual risk assessment. This phase is primarily a review phase, where the CEMT takes the data and relationships defined in the earlier phases and builds new views of those data to support the assessment and analysis by stakeholders.

The first such view is an attack tree, an example of which is shown in Figure 8. These attack trees trawl the threat flowcharts defined in the Threat Modelling phase and present this information as a directed acyclic graph where each node is a step taken by the threat actor. This results in an attack tree where every branch of the tree represents a single threat vector, start to finish. The purpose of this diagram is to provide a summary view of the entire attack surface, and allows stakeholders to review the threat model to ensure that there is sufficient coverage of important threat vectors.

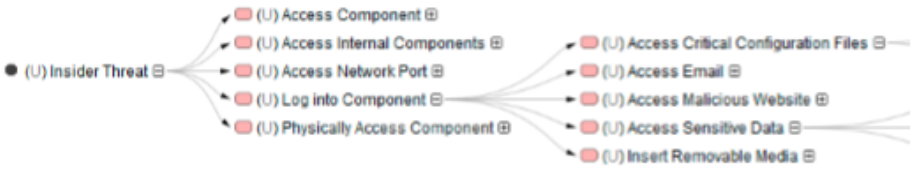


Figure 8. Partial CEMT attack tree—forward diagram from the CEMT sample model.

In addition to the attack trees, the CEMT also builds bowtie diagrams that allow the modeler to select a node in the attack tree, and the bowtie diagrams display the security controls that are relevant to that node, grouped into mitigating controls and detection controls, colour-coded by the implementation state of that control based on the audit results input into the CEMT. In conjunction with the attack trees, this allows stakeholders to query each node to determine what the system is doing to address that step in the threat vector.

As an example, the assessment stakeholders may focus on nodes that are common to a number of branches on the attack tree, as implementing additional controls at those nodes are likely to yield bigger benefits as they will simultaneously mitigate multiple threat

vectors. Equally, a focus may be placed on particularly short branches of the attack tree, as these represent situations where there is not much natural defence-in-depth, which means that the mitigations on the nodes in that branch are often critically important. Ultimately, the purpose of this assessment is to provide a qualitative review of the entire attack surface, ensure that critical nodes are adequately mitigated, and identify potentially high-risk threat vectors for further analysis via quantitative simulation.

Once the high-risk threat vectors have been identified, a detailed analysis can be performed on those threat vectors using the CEMT parametric risk diagram. These diagrams are created by the CEMT for the high-risk threat vectors identified during the analysis of the attack trees. Each node in the threat vector is added to the parametric diagram along with the associated detection action and chains them together to set up a parametric simulation that calculates a residual probability and detection probability for the threat vector. A partial extract of one of these diagrams is shown in Figure 9, which shows two nodes of the threat vector and also depicts the related control effectiveness values and the linked controls, colour-coded by their implementation state.

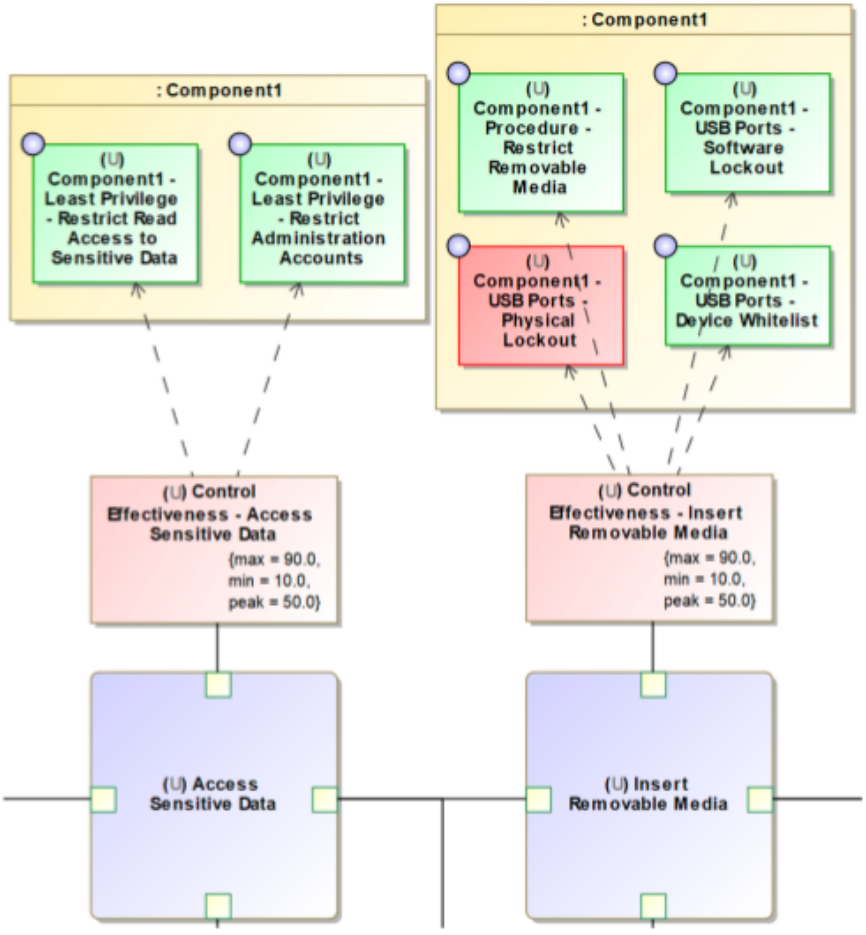


Figure 9. Partial CEMT parametric risk diagram from the CEMT sample model.

Importantly, Figure 9 highlights the control effectiveness value, which is a critical input that needs to be provided to the CEMT during this step. The control effectiveness value is the percentage of attempts of the action that would be prevented by the controls that are in place. This value is likely to be subjective, so the modeler can input a value range to account for the uncertainty in the evaluation of this input value. This function defaults

to a triangular distribution with a minimum, maximum, and peak value. The process is repeated for the associated detection actions, and once these values have all been set, the CEMT can run a Monte-Carlo simulation and produce a histogram of the resulting residual risk probability and detection probability. An example of these histograms is shown in Figure 10.

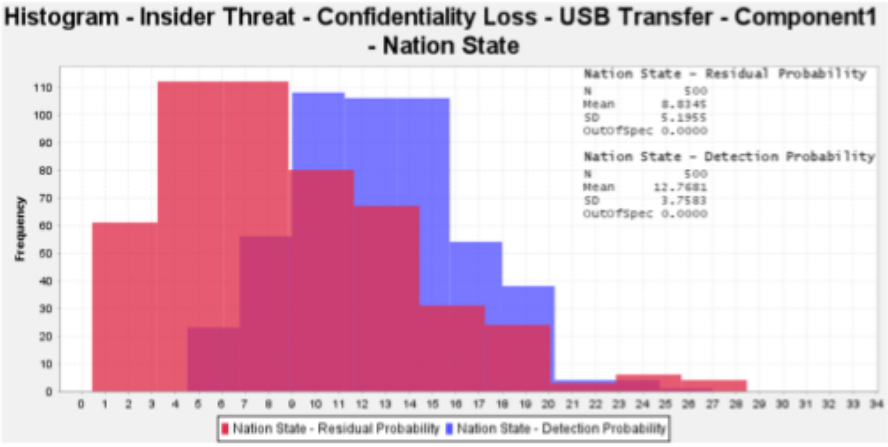


Figure 10. Simulation histogram for the CEMT parametric risk diagram.

The residual risk is only realised when all of the steps are successfully completed by the adversary, so the calculation is a simple multiplication of the probabilities of each step being completed successfully, that is:

$$P_{Residual} = P_{Initial} \times P_{1_T} \times P_{2_T} \times \ldots \times P_{x_T} \quad (1)$$

where $P_{Initial}$ is the probability of the attack being attempted by an adversary, as set by the modeler, and P_{x_T} is the probability of the individual threat node being successfully completed by the attacker. P_{x_T} is calculated from the multiplication of two values, E_{x_T}, which is derived from the control effectiveness value input by the modeler for that threat node, and the difficulty input, D_{x_T}, which is a user-configurable mapping between the difficulty values assigned to each threat node in the Threat Modelling phase and a quantitative probability that they would be able to complete the attack step:

$$P_{x_T} = (E_{x_T} \times D_{x_T}) \quad (2)$$

E_{x_T} is defined as the percentage of attacks that would not be prevented by the controls that have been implemented on that threat node and can be based on either historical data or the expert judgement of the modelers.

D_{x_T} is defined as the percentage of attempted attacks that would be successful in isolation of any controls. D_{x_T} represents the difficulty of the step irrespective of any mitigating controls and is driven by the capability of the threat actor. As an example, a cyber payload delivered via an electronic attack against a signal processor would be inherently more difficult than a simpler node such as inserting a USB device, and this difference is captured in the difficult input term.

This probability calculation treats each threat node as an independent event, and where this does not hold true, for example, if a control set is reused later in a particular risk chain, then the modeler would need to account for this in the effectiveness value for the second appearance of the control in that chain. If the attack progresses because the control was ineffective the first time it appeared, it is unlikely to be effective the second time the attacker encounters it later in the chain, and the modeler must reduce the effectiveness in the later node to ensure accuracy.

Detection, on the other hand, can occur at any step, and a single successful detection node would result in the attack being detected. Thus, the overall detection probability is calculated as a probability summation:

$$P_{Detection} = [P_{Initial} \times P_{1_D}] + [P_2 \times P_{2_D}] + \ldots + [P_x \times P_{x_D}] \quad (3)$$

where P_{1_D} is the probability of the individual detection node being successful and is calculated as the multiplication of two values, E_{x_D}, which is derived from the control effectiveness value input by the modeler for that detection node, and the avoidance input, A_{x_D}, which is a user-configurable mapping between the capability of the threat actor and a quantitative probability that they would avoid the detecting controls:

$$P_{x_D} = (E_{x_D} \times A_{x_D}) \quad (4)$$

E_{x_D} is defined as the percentage of attempts that would be detected by the controls that have been implemented on that detection node and can be based on either historical data or the expert judgement of the modelers. A_{x_D} is defined as the percentage of potential detections that would be avoided by the attacker, based on the capability of the attacker to avoid detection, and can be based on either historical data or the expert judgement of the modelers. P_x is the probability that the detection node will be triggered, which occurs if the previous threat node was successfully completed, therefore:

$$P_x = P_{Initial} \times P_{1_T} \times P_{2_T} \times \ldots \times P_{(x-1)_T} \quad (5)$$

Finally, the CEMT builds a summary risk table using these simulation values, along with the steps required to complete threat vectors, a list of the participating assets, and a list of the relevant mitigation mechanisms. The assessment stakeholders will then map that information to the qualitative likelihood values used within their particular organisation to assess the consequence of the threat vector succeeding and determine the final risk rating for that particular threat vector. Once all critical threat vectors have been assessed, this summary risk table will be populated with a risk assessment for each of the high priority threats to the system, completing the CEMT process.

2.4. Achieving Threat-Based Cybersecurity Engineering

The CEMT approach effectively implements the threat-based cybersecurity engineering concepts outlined earlier in this paper. The CEMT threat model is derived directly from the system context and can be informed by the Cyber Threat Intelligence function within an organisation, who can work collaboratively with the modelers to determine the most appropriate misuse cases based on known threat actors. Threat emulation teams can be engaged to help analyse the system design through the lens of those misuse cases, provide valuable insight into the feasibility of certain attacks, and validate the specific steps that would need to be taken to complete those attacks, as articulated by the mal-activity diagrams.

Best-practice control frameworks are leveraged to determine the set of controls that are tied to the specific steps of the threat model, but the controls from those best-practice frameworks are only incorporated where they are addressing a specific threat-emulation-validated threat action derived from the threat-intelligence-informed misuse cases.

The overall output of the CEMT approach is a set of mitigating and detecting controls that can be implemented to manage risk within the system, and a calculated residual risk based on the implementation state of those mitigating and detecting controls. This achieves all of the outputs depicted in Figure 2, with the exception of the recovery controls. The incorporation of recovery controls into the CEMT approach is an intended extension to the process, and this is discussed further in Section 4.2 of this paper.

This is all achieved within an assessment methodology based on graphical representations of the cybersecurity risks in the system, allowing for transparent collaboration between disparate stakeholders. The underlying structured datasets can also be lever-

aged, much like model-based system design, to increase the efficiency of the modelling by improving the reusability, maintainability, and speed of the assessment using automation.

2.5. *Efficiency through Automation*

The ability to increase the efficiency of complex system development is seen as one of the potential benefits of the model-based systems engineering approach [52], and this ability to automate in order to minimise the engineering impact of iterative change to the design is a key component of the application of agile methods into system engineering [53]. It is important that the cybersecurity analysis of these complex systems takes advantage of the potential of automation during model-based risk assessments to ensure that it supports, rather than becomes a roadblock to, the need for efficient iteration of the system design. There are three primary methods of automation supported by the CEMT:

- Automated update of complex drawings and simulations to ensure that changes to the design or threat environment can be incorporated efficiently into the threat model;
- Automated model validation to ensure that basic review tasks are automated, allowing expert reviewers to focus on the actual threat assessment component;
- Automated documentation to ensure that the process of creating enduring design artefacts is efficient and accurate.

The automation of complex drawings and simulations in the CEMT is achieved through a combination of bespoke relationship expressions that leverage the capabilities inherent in the underlying digital engineering tool and purpose-built JavaScript API calls that extend these inherent capabilities to achieve the desired views in the CEMT.

The first example of this is the development of the attack trees and bow-tie diagrams in the Risk Assessment phase. While attack trees are not particularly novel and have been used in security analysis since at least 1994 [54], one of their limitations is that they are primarily static [13] and are difficult to update once a significant level of complexity or interconnectedness is reached. The CEMT deliberately uses mal-activity diagrams to allow the modeler to express the threat vectors in a manner that maximises the efficiency of inputting the data, but then automatically generates an attack tree using queries of the structured information within those mal-activity diagrams, which provides the same information in a format that is significantly easier to review. This allows for changes to be made to the mal-activity diagram as the system design matures or the threat environment changes, and the CEMT will automatically keep the attack trees up to date.

This drawing automation is extended through the use of JavaScript API calls to generate new diagrams from existing data in the model where the in-built query language does not sufficiently support the necessary level of automation. The generation of parametric diagrams and the associated simulations in the Risk Assessment phase is automated in this way, allowing a modeler to select a threat vector in the attack tree and generate the simulation in an automated fashion. Similarly, the simulations can be run and the results saved into the model through the automation macros provided in the CEMT.

One of the limitations of using JavaScript API in this way is that the generated diagrams are not kept up to date when the source data in the mal-activity diagrams are changed. To address this limitation, the second type of automation—automated model validation—is also used in the CEMT. This automation leverages the active validation functionality in the underlying digital engineering software to provide real-time validation of a model against a defined set of constraints. The CEMT provides several active validation constraints that will be highlighted to the modeler when one of the diagrams built through the JavaScript API is out-of-date due to a change in the source data, allowing the modeler to rerun the automation macro to resolve the inconsistency. In addition to this, a number of active validation constraints are included that provide guidance to the modeler as they are building out the model to ensure the completeness of the model. This real-time assistance to the modeler helps to resolve issues as early in the modelling process as possible while also removing the burden of identifying these errors from a manual peer review team.

The final type of automation used in the CEMT is automated document generation. While one of the primary purposes of digital engineering is to move away from document-based engineering analysis [55–57], there remains a need to provide a snapshot of the model at particular points in time for the purposes of record-keeping, archiving, and sharing information with stakeholders who may not have access to digital engineering tools. Using a system safety engineering analogy, the threat model can be thought of as the cybersecurity equivalent of the safety case [58], but there is still a practical need for a snapshot of the threat model much like there is a practical need for a safety case report [59]. To support this, the CEMT provides an automated export of the output of the threat model into a Microsoft Word document that summarises the scope of the misuse cases that have been evaluated and presents the implementation status of the control baseline as well as the results of the risk assessment on the critical threat vectors. This provides an artefact that can be quickly and easily generated and regenerated, which can then be distributed into an organisation's existing risk governance processes for endorsement and acceptance.

3. Results

3.1. Face Validity Trial Setup

Initial research was conducted on the CEMT to assess the face validity of the proposed approach. This investigation assessed whether the approach seemed promising to relevant experts working in cybersecurity evaluations of complicated systems to inform continued future work on our approach and associated tools. Presentations, tutorials, and workshops on the CEMT process and the worked example described in Section 2.3 were provided to many cybersecurity engineers, systems engineers, and project managers across the communities of a major Australian government department and the Australian power industry during 2022. Participants were offered the opportunity to complete a short survey to capture their impressions of the approach with respect to the effectiveness, efficiency, and expected benefits.

3.2. Face Validity Trial Data Collection and Setup

A total of 20 responses were received (n = 20). Fourteen respondents identified themselves as having a background as a cybersecurity subject matter expert, with an average experience of 7.8 years. Sixteen respondents identified themselves as having worked with cybersecurity risk assessments in some capacity, with an average experience of 5.5 years.

The breakdown of the organisation type of the respondents consisted of 40% coming from public sector Defence and 60% coming from private industry. There was also an even spread of disciplines represented in the survey. A total of 30% of the respondents described themselves as fulfilling a governance role in the development of cyber-physical systems, 30% identified themselves as fulfilling a management role, 30% identified themselves as fulfilling a technical specialist role, and 10% responding with other.

However, the survey respondents did see themselves as having a technical dimension to the main area of expertise, as outlined in Figure 11. A total of 70% of those surveyed identified their main area of expertise as either IT/Computing or engineering, with only 10 per cent selecting management as their main area of expertise. Figure 12 highlights the spread of expected roles in the CEMT process amongst the survey responses. These results showed a reasonable spread between the different roles, with no one role dominating the responses.

Overall, the demographics of the survey respondents were balanced between different specialisations within the process of cybersecurity risk assessments. The survey participants had a strong level of experience in cybersecurity, both generally and in terms of risk assessment. The bias towards a technical engineering/IT background is to be expected as the respondents were selected from attendees at professional presentations and conference tutorials in the development of complex cyber-physical systems, which would naturally select participants with that type of background experience. Ultimately, this cohort presents

a reasonable foundation from which to draw conclusions about the face validity of the CEMT process.

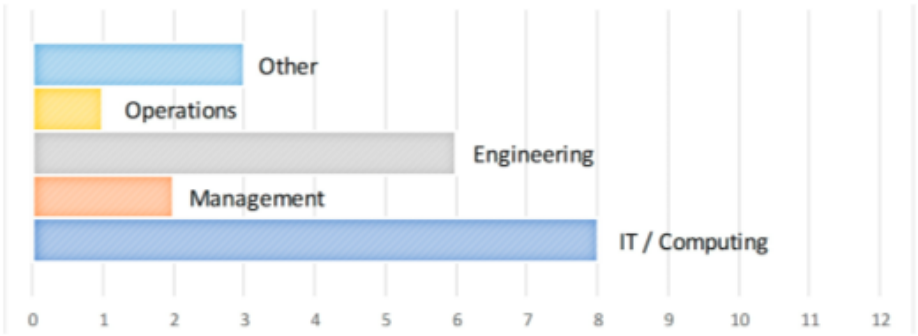


Figure 11. Responses to the question: "What is your main area of expertise?" Bars indicate the number of respondents, out of the 20 respondents, that provided that particular response.

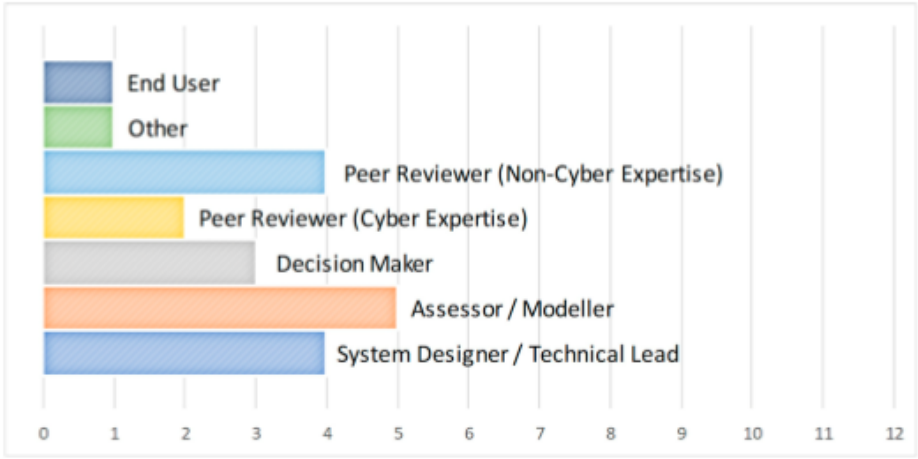


Figure 12. Responses to the question: "What option best describes your expected role in a cyberworthiness assessment process?" Bars indicate the number of respondents, out of the 20 respondents, that provided that particular response.

The survey responses are included in Table 1, with the summary results graphed in Figure 13. The responses were broadly positive, with most responses supporting the expected benefits of a cyber assessment using the CEMT.

These results generally support the claims that the approach can produce contextually relevant, collaborative assessments that provide an improved understanding of a system's cyberworthiness. However, they also show concerns about the difficulty of producing the assessment, its overall accuracy, and the model's ability to address cyberworthiness. Nevertheless, the overall response to the approach in the survey was positive, with 75% of respondents giving the approach an overall rating of either excellent or very good, as depicted in Figure 14.

These face validity findings are not being used as scientific evidence to confirm the technical validity of the approach, but rather as an indicator of the expected benefits of the approach, informing the appropriateness of future research to fully trial its technical validity.

Table 1. Responses to the survey questions. Table cells indicate the percentage (%) of the 20 respondents that selected the particular response to each question.

	Survey Question	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Q1	The CEMT produces risk assessments that are tailored to the context in which the system operates	0	0	15	50	35
Q2	Cyberworthiness assessments are simple to produce using the CEMT	5	0	40	35	20
Q3	The CEMT is an effective use of time	0	0	30	25	45
Q4	The CEMT process is intuitive	0	5	25	45	25
Q5	The CEMT encourages stakeholders to work collaboratively to determine the residual risk level	0	0	10	35	55
Q6	The CEMT clearly identifies which security controls are important to the system	0	0	5	55	40
Q7	The CEMT produces transparent cyberworthiness assessments	0	5	10	40	45
Q8	The CEMT facilitates informed decision making with respect to the identified cybersecurity risks	0	0	5	50	45
Q9	The CEMT produces cyberworthiness assessments that have ongoing value through the future phases of the capability life cycle	0	0	10	40	50
Q10	The CEMT would improve my understanding of the cyberworthiness of a system	0	0	10	20	70
Q11	The CEMT produces accurate assessments of a system's cyberworthiness	0	10	20	35	35
Q12	The CEMT facilitates the engagement of stakeholders and the provision of meaningful input from those stakeholders into a cyberworthiness assessment	0	0	20	40	40
Q13	The cyberworthiness assessments produced by the CEMT are sufficiently detailed	0	5	20	30	45
Q14	The CEMT identifies the relative impact of security controls with respect to the cyberworthiness of the system	0	5	15	40	40
Q15	The CEMT is not overly dependent on the subjective opinion of subject matter experts	0	0	30	50	20
Q16	The CEMT provides sufficient information to allow decision makers to be accountable for their decisions	0	10	15	35	40
Q17	The CEMT clearly highlights the areas of greatest cyber risk to the system	0	0	15	35	50
Q18	The CEMT adds value to a system and/or project	0	0	5	35	60
Q19	The CEMT provides a complete and comprehensive approach to determining cyberworthiness	5	10	10	50	25
Q20	The CEMT is an improvement over existing cyberworthiness assessment processes	0	5	10	20	65

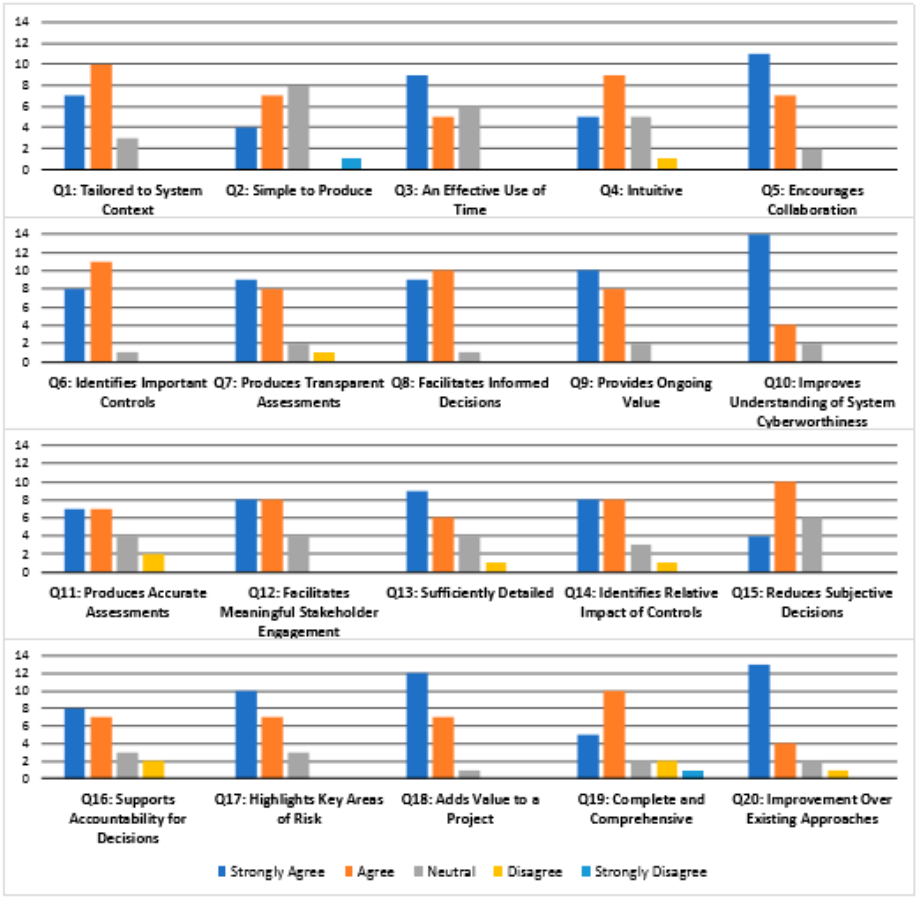


Figure 13. Summary results of the CEMT face validity survey. Bars indicate the number of respondents, out of the 20 respondents, that provided that particular response.

Figure 14. Overall rating of the threat-based cybersecurity engineering approach using the CEMT, based on responses to the face validity survey. Number in each pie segment indicates the number of respondents, out of the 20 respondents, that provided that particular response.

4. Discussion

The CEMT was able to meet most of the expectations of the surveyed users concerning outcomes, where they broadly welcome threat-based cybersecurity engineering using graph-based approaches for cybersecurity analysis created using model-based systems engineering techniques. The tool promises a repeatable and consistent method for producing (1) abuse case graphs, referred to as misuse case diagrams in the CEMT, and (2) insecurity flow graphs, referred to as mal-activity diagrams in the CEMT, which contain the necessary model-based attributes that allow for the dynamic generation of data views. Surveyed users welcomed the feature for static graph techniques such as attack trees to be generated in real-time to help ensure alignment between the attack trees and insecurity flow graphs and to simplify the maintenance of these complementary approaches.

The CEMT also provides a promising framework for prioritising the importance of different mitigation techniques by explicitly tracing mitigation techniques to threat vectors and then tying those threat vectors to a quantitative analysis of the risk likelihood. This feature aims to allow cyber analysts to rapidly reassess different configurations of the design that have different mitigation implementation states and recalculate the likelihood of critical threat vectors. A secondary advantage of using common model-based systems engineering tools as the basis of the CEMT is that it can also be integrated into design models of a system that has been developed using these model-based approaches.

There are still some limitations of the CEMT that were highlighted in the validity workshops and when compared to the generic threat-based cybersecurity engineering model presented at the start of this paper. While the mitigating and detecting controls were a clear output of the CEMT alongside the residual risk ratings, the tool does not currently derive recovery controls. Deriving these controls within the toolkit is an avenue for future development of the tool and could likely be achieved by creating a set of activity diagrams that address recovery from the identified high-risk threat vectors. These diagrams would articulate a set of actions an operations team would need to take to recover the system in the event of that risk being realised, and the recovery controls that would make the team's efforts possible or more effective would be traced to each step in that recovery process. The level of implementation of these controls could then mitigate the consequence of the risk, as the ability and responsiveness of a recovery action would directly mitigate the consequence of a risk being realised.

The results of our face validity trial into the CEMT show promise for the ability of the toolkit to meet the needs of assessing the cyberworthiness of a system using a threat-based cybersecurity engineering approach. The overall rating of the approach amongst the respondents was overwhelmingly positive, with 40% of responses rating it as excellent and a further 35% rating it as very good.

The surveys also confirmed that the respondents believed that the CEMT could address the three primary challenges that the current compliance-based risk assessments face, that is:

- Appropriateness of the assessed controls to the system being assessed, as demonstrated by the responses to Question 1;
- Prioritisation of controls, as demonstrated by the responses to Questions 6 and 14;
- Ability for non-expert decision makers to understand the assessment, as demonstrated by Questions 7, 8, and 17.

The existing compliance-based assessments such as the Australian Government PSPF [1] and NIST RMF [2] provide a consistent list of best-practice controls that are applied against a system regardless of its operating environment, while the threat-based approach of the CEMT starts from the system context and its operational environment to identify threat vectors and then link best-practice security controls that are relevant to those threat vectors. The developed methodology inherently constrains the scope of control assessments to the context of the system being assessed.

As these controls are then explicitly traced to the threat vectors, the importance of controls can be analysed based on the number of threat vectors that each control mitigates

as well as the criticality of those threat vectors in terms of the likelihood of the threat vector being successful and the consequence of the impact of the threat vector. This architectural view provides a basis for having informed discussions with decision makers as to the importance of implementing a particular control, or not.

These discussions with decision makers are further supported by the CEMT due to its graphical nature and explicit modelling. Analysts produce step-by-step flowcharts of each possible threat vector, and control effectiveness is described in terms of reducing the likelihood of a single step being performed by an attacker. This provides a lattice for robust conversations around assumptions and exclusions in the modelling, removing much of the opaque expert judgement inherent in high-level cyber risk assessments.

The results of the face validity survey also confirmed that some concerns around the CEMT approach remain, particularly in terms of the simplicity to produce and the overall efficiency of the project. While significant steps were made between the early prototypes and the CEMT in terms of usability and simplicity, the survey responses indicate that at least some concern remains about these facets of the CEMT approach.

4.1. Significance

In 2015, Nguyen et al. [60] provided a comprehensive review of the literature associated with using model-based techniques for the security assessment of cyber-physical systems. This state of research was revisited by Geissman and Bodden [61] in 2020. They found several approaches and methodologies in the literature that leveraged model-based techniques including SysML-based approaches to address the security assessment of complex cyber-physical systems.

Many of these existing approaches such as the cyber security requirements methodology (CSRM) [62] and Larsen et al. [63] are focused on the definition of security requirements and provide a methodology for managing the implementation of those security requirements through the systems design life cycle. Our CEMT approach distinguishes itself from these approaches by being explicitly threat focused, and uses this threat analysis to derive the relevant security properties within the system.

Several other methodologies including SAVE [64], Navas et al. [65], and Geissman et al. [66] are also threat focused and utilise similar threat modelling techniques such as misuse cases to the CEMT. However, these approaches provide a high-level threat model that interacts with a detailed white-box model of the system. The CEMT extends the misuse case approach with detailed mal-activity diagrams that create a white-box threat model that interacts with either a black-box or white-box system model. This feature enables modellers to create detailed threat models to explain the cybersecurity threat to non-cybersecurity experts including non-technical decision makers.

Some emerging methodologies such as MBSESec [67] also provide more detailed modelling of adversary behaviour in the form of activity diagrams. However, the novelty provided by the CEMT in comparison to these lies in the explicit control traceability to the detailed threat model and the use of specific SysML profiles that support the rapid creation of summary visualisations such as attack trees and bow-tie diagrams. While this traceability and visualisation does not necessarily produce a better or more accurate threat model, it does enhance the ability of non-specialists to interact with and critically review the threat model to make more informed risk decisions.

The threat focus of the CEMT is also the primary point of difference between it and the other existing extension profiles that address security assessments in model-based languages. For example, UAF [68], UMLSec [69], and SysML-Sec [70] all provide taxonomies and extension profiles that focus on the allocation of security requirements and the linkages of assets, vulnerabilities, and mitigations. The primary point of difference in the CEMT is that security requirements are derived from the threat modelling process itself, and that the allocation of mitigations to assets is explicitly performed via common relationships to the threat model.

To illustrate the significance of the CEMT approach, five key attributes that contribute to a comprehensive cyberworthiness evaluation were defined:

- Extended Model-Based Taxonomy—an extension of an open model-based systems engineering language such as UML or SysML; this is provided to facilitate a model-based approach;
- Threat Focused—the threats to the system, rather than a best-practice control baseline or asset hierarchy, is used as the focal point of the assessment;
- Detailed Adversary Modelling—the actions of the adversary are modelled in detail, facilitating a precise discussion and review of any threat analysis;
- Visualisation and Simulation of Threat—detailed adversary modelling is expressed in simplified graphs such as attack trees, and branches of those graphs can be simulated quantitatively;
- Explicit Traceability to Threats—derived security controls are directly traceable to adversary actions, facilitating discussion and review of the importance of each control in terms of the malicious action it mitigates.

Table 2 provides a summary comparison of these published model-based security assessment frameworks to the CEMT across these five key attributes. Figure 15 provides a visual depiction of the level to which these five key attributes are served by these existing frameworks, with green representing those attributes that are well-served by existing techniques, yellow indicating attributes that are served by some of the existing techniques, and red indicating those attributes that are not currently found in existing techniques.

Table 2. Comparison of model-based security assessments along five key attributes.

	Model-Based Security Assessment Approach	Extended Model-Based Taxonomy	Threat Focused	Detailed Adversary Modelling	Visualisation and Simulation of Threats	Explicit Traceability to Threats
1	CSRM [62]	Y	N	N	N	N
2	Larsen et al. [63]	Y	N	N	N	N
3	SAVE [64]	Y	Y	N	N	N
4	Navas et al. [65]	Y	Y	N	N	N
5	Geissman et al. [66]	Y	Y	N	N	N
6	MBSESec [67]	Y	Y	Y	N	N
7	UAF [68]	Y	N	N	N	N
8	UMLSec [69]	Y	N	N	N	N
9	SysML-Sec [70]	Y	N	N	N	N
10	CEMT	Y	Y	Y	Y	Y

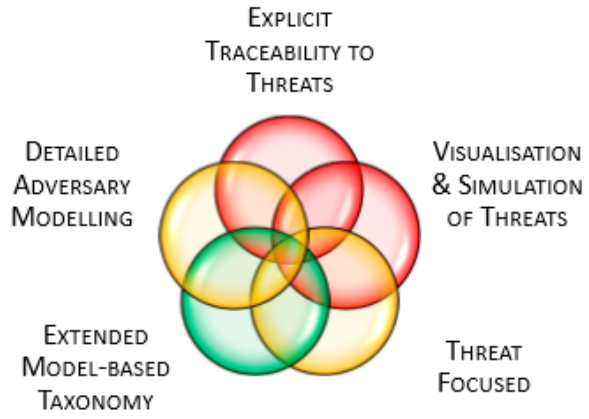


Figure 15. Significance of the CEMT approach across five key attributes of model-based security assessments.

4.2. Future Work

While the results to date have been promising, there are still a number of avenues for future work, which can be broadly grouped into further refinement of the toolkit, further expansion of the toolkit, and further case study trial implementation of the toolkit.

Refinement of the toolkit involves continuing to address the latest concerns around usability, simplicity, and efficiency. This refinement could be achieved by developing further automation within the toolkit to increase the user experience of creating models, producing additional training materials to help with the initial learning process or through the development of reusable libraries of common threat vectors, similar to the generic case study examples.

Expansion of the toolkit could involve the development of new modelling steps or diagram views to increase the effectiveness of the toolkit. This effort would likely include the expansion of the toolkit to address recovery from a successful cyber-attack through the development of response activity diagrams from which are derived a set of relevant recovery controls. It could also include the development of additional summary diagrams to aid in the explanation of the cyber risk to different audiences.

The primary future work, however, is through further implementation of the CEMT approach. The face validity trials provide a view of the expected benefits of the toolkit, but to assess the overall technical validity of the approach, further implementation on real-world systems must be completed and evaluated. These implementation trials are currently underway, with scope for wider trials and research into other industries and countries. We look forward to reporting on the results once the initial trials are complete.

5. Conclusions

The CEMT provides an example of a model-based technique that can support decision making around the acceptability of cybersecurity risk in complicated engineered systems. The cybersecurity assessment methodology is unique in utilising model-based systems engineering techniques to develop a detailed threat model of adversary behaviour that defines explicit traceability between controls and threats and provides a visualisation and simulation of any resultant risks. We have provided preliminary results based on a face validity trial that confirms that there is significant potential for the methodology and toolkit to positively impact the effectiveness of this decision-making process. This face validity has enabled detailed implementation trials to proceed. The widespread availability of the CEMT and its face validity should enable organisations beyond the Australian microcosm to trial this framework and toolset: we welcome wider use and feedback to keep ahead of the advanced persistent threat.

Author Contributions: Conceptualisation, S.F. and K.J.; Methodology, S.F.; Software, S.F.; Investigation, S.F.; Writing—original draft preparation, S.F.; Writing—review and editing, K.J. and S.M.; Visualisation, S.F.; Supervision, K.J. and S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The models and tools presented in the study are openly available in GitHub at https://github.com/stuartfowler/CEMT/ (accessed on 25 April 2024). The raw survey data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: This research is supported by an Australian Government Research Training Program (RTP) Scholarship.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Australian Government—Department of Home Affairs, Protective Security Policy Framework. Available online: https://www.protectivesecurity.gov.au (accessed on 25 April 2024).
2. National Institute of Standards and Technology (NIST) Computer Security Resource Center (CSRC), NIST Risk Management Framework (RMF). Available online: https://csrc.nist.gov/projects/risk-management/about-rmf (accessed on 25 April 2024).
3. Australian Government—Australian Signals Directorate, Information Security Manual (ISM). Available online: https://www.cyber.gov.au/resources-business-and-government/essential-cyber-security/ism (accessed on 25 April 2024).
4. National Institute of Standards and Technology (NIST) Computer Security Resource Center (CSRC), NIST Special Publication 800-53 Rev. 5: Security and Privacy Controls for Information Systems and Organizations. Available online: https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final (accessed on 25 April 2024).
5. U.S. Department of Navy; Cyber Strategy, November 2023. Available online: https://dvidshub.net/r/irstzr (accessed on 25 April 2024).
6. Australian Government—Australian Signals Directorate, System Security Plan Annex Template (March 2024). Available online: https://www.cyber.gov.au/sites/default/files/2024-03/System%20Security%20Plan%20Annex%20Template%20(March%202024).xlsx (accessed on 25 April 2024).
7. National Institute of Standards and Technology (NIST) Computer Security Resource Center (CSRC), Control Catalog (spreadsheet). Available online: https://csrc.nist.gov/files/pubs/sp/800/53/r5/upd1/final/docs/sp800-53r5-control-catalog.xlsx (accessed on 25 April 2024).
8. National Institute of Standards and Technology (NIST), OSCAL: The Open Security Controls Assessment Language. Available online: https://pages.nist.gov/OSCAL/ (accessed on 25 April 2024).
9. MITRE ATT&CK Framework. Available online: https://attack.mitre.org/ (accessed on 25 April 2024).
10. The Department of Defense Cyber Table Top Guide, Version 2, 16 September 2021. Available online: https://www.cto.mil/wp-content/uploads/2023/06/DoD-Cyber-Table-Top-Guide-v2-2021.pdf (accessed on 25 April 2024).
11. Monroe, M.; Olinger, J. Mission-Based Risk Assessment Process for Cyber (MRAP-C). *ITEA J. Test Eval.* **2020**, *41*, 229–232.
12. Kuzio de Naray, R.; Buytendyk, A.M. *Analysis of Mission Based Cyber Risk Assessments (MBCRAs) Usage in DoD's Cyber Test and Evaluation*; Institute for Defense Analyses: Alexandria, VA, USA, 2022; IDA Publication P-33109.
13. Kordy, B.; Pièvre-Cambacédès, L.; Schweitzer, P.P. DAG-based attack and defense modeling: Don't miss the forest for the attack trees. *Comput. Sci. Rev.* **2014**, *13–14*, 1–38. [CrossRef]
14. Weiss, J.D. A system security engineering process. In Proceedings of the 14th Annual NCSC/NIST National Computer Security Conference, Washington, DC, USA, 1–4 October 1991.
15. Schneier, B. Attack trees: Modeling security threats. *Dr Dobb's J. Softw. Tools* **1999**, *12–24*, 21–29. Available online: https://www.schneier.com/academic/archives/1999/12/attack_trees.html (accessed on 25 April 2024).
16. Paul, S.; Vignon-Davillier, R. Unifying traditional risk assessment approaches with attack trees. *J. Inf. Secur. Appl.* **2014**, *19*, 165–181. [CrossRef]
17. Kordy, B.; Pouly, M.; Schweitzer, P. Probabilistic reasoning with graphical security models. *Inf. Sci.* **2016**, *342*, 111–131. [CrossRef]
18. Gribaudo, M.; Iacono, M.; Marrone, S. Exploiting Bayesian Networks for the analysis of combined Attack Trees. *Electron. Notes Theor. Comput. Sci.* **2015**, *310*, 91–111. [CrossRef]
19. Holm, H.; Korman, M.; Ekstedt, M. A Bayesian network model for likelihood estimations of acquirement of critical software vulnerabilities and exploits. *Inf. Softw. Technol.* **2015**, *58*, 304–318. [CrossRef]
20. Moskowitz, I.; Kang, M. An insecurity flow model. In Proceedings of the 1997 Workshop on New Security Paradigms, Cumbria, UK, 23–26 September 1997; pp. 61–74.
21. McDermott, J.; Fox, C. Using abuse case models for security requirements analysis. In Proceedings of the 15th Annual Computer Security Applications Conference, Phoenix, AZ, USA, 6–10 December 1999; pp. 55–64.
22. Sindre, G.; Opdahl, A.L. Eliciting security requirements with misuse cases. *Requir. Eng.* **2004**, *10*, 34–44. [CrossRef]
23. Karpati, P.; Sindre, G.; Opdahl, A.L. Visualizing cyber attacks with misuse case maps. In *Requirements Engineering: Foundation for Software Quality*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 262–275.
24. Abdulrazeg, A.; Norwawi, N.; Basir, N. Security metrics to improve misuse case model. In Proceedings of the 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensics, Kuala Lumpur, Malaysia, 26–28 June 2012.
25. Saleh, F.; El-Attar, M. A scientific evaluation of the misuse case diagrams visual syntax. *Inf. Softw. Technol.* **2015**, *66*, 73–96. [CrossRef]
26. Mai, P.; Goknil, A.; Shar, L.; Pastore, F.; Briand, L.C.; Shaame, S. Modeling Security and Privacy Requirements: A Use Case-Driven Approach. *Inf. Softw. Technol.* **2018**, *100*, 165–182. [CrossRef]
27. Matuleviaius, R. *Fundamentals of Secure System Modelling*; Springer International Publishing: Cham, Switzerland, 2017; pp. 93–115.
28. Sindre, G. Mal-activity diagrams for capturing attacks on business processes. In *Requirements Engineering: Foundation for Software Quality*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 355–366.
29. Opdahl, A.; Sindre, G. Experimental comparison of attack trees and misuse cases for security threat identification. *Inf. Softw. Technol.* **2009**, *51*, 916. [CrossRef]
30. Karpati, P.; Redda, Y.; Opdahl, A.; Sindre, G. Comparing attack trees and misuse cases in an industrial setting. *Inf. Softw. Technol.* **2014**, *56*, 294. [CrossRef]

31. Tondel, I.A.; Jensen, J.; Rostad, L. Combining Misuse Cases with Attack Trees and Security Activity Models. In Proceedings of the 2010 International Conference on Availability, Reliability and Security, Krakow, Poland, 15–18 February 2010; pp. 438–445.
32. Meland, P.H.; Tondel, I.A.; Jensen, J. Idea: Reusability of threat models—Two approaches with an experimental evaluation. In *Engineering Secure Software and Systems*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 114–122.
33. Purton, L.; Kourousis, K. Military Airworthiness Management Frameworks: A Critical Review. *Procedia Eng.* **2014**, *80*, 545–564. [CrossRef]
34. Mo, J.P.T.; Downey, K. System Design for Transitional Aircraft Support. *Int. J. Eng. Bus. Manag.* **2014**, *6*, 45–56. [CrossRef]
35. Hodge, R.J.; Craig, S.; Bradley, J.M.; Keating, C.B. Systems Engineering and Complex Systems Governance—Lessons for Better Integration. *INCOSE Int. Symp.* **2019**, *29*, 421–433. [CrossRef]
36. Simmonds, S.; Cook, S.C. Use of the Goal Structuring Notation to Argue Technical Integrity. *INCOSE Int. Symp.* **2017**, *27*, 826–841. [CrossRef]
37. United States Government Accountability Office. Weapon Systems Cybersecurity: DOD just Beginning to Grapple with Scale of Vulnerabilities. *GAO-19-129*. 2018. Available online: https://www.gao.gov/products/gao-19-128 (accessed on 15 June 2024).
38. Joiner, K.F.; Tutty, M.G. A tale of two allied defence departments: New assurance initiatives for managing increasing system complexity, interconnectedness and vulnerability. *Aust. J. Multi-Discip. Eng.* **2018**, *14*, 4–25. [CrossRef]
39. Joiner, K.F. How Australia can catch up to U.S. cyber resilience by understanding that cyber survivability test and evaluation drives defense investment. *Inf. Secur. J. A Glob. Perspect.* **2017**, *26*, 74–84. [CrossRef]
40. Thompson, M. Towards Mature ADF Information Warfare—Four Years of Growth. *Defence Connect Multi-Domain*. 2020. Available online: https://www.defenceconnect.com.au/supplements/multi-domain-2 (accessed on 15 June 2024).
41. Fowler, S.; Sweetman, C.; Ravindran, S.; Joiner, K.F.; Sitnikova, E. Developing cyber-security policies that penetrate Australian defence acquisitions. *Aust. Def. Force J.* **2017**, *102*, 17–26.
42. Australian Senate. Budget Hearings on Foreign Affairs Defence and Trade, Testimony by Vice Admiral Griggs, Major General Thompson and Minister of Defence (29 May, 2033–2035 hours). 2018. Available online: https://parlview.aph.gov.au/mediaPlayer.php?videoID=399539timestamp3:19:43 (accessed on 15 June 2024).
43. Australian Government. *ADF Cyberworthiness Governance Framework*; Australian Government: Canberra, Australia, 2020.
44. Australian Government. Defence Seaworthiness Management System Manual. 2018. Available online: https://www.defence.gov.au/sites/default/files/2021-01/SeaworthinessMgmtSystemManual.pdf (accessed on 15 June 2024).
45. Allen, M.S.; Robson, D.A.; Iliescu, D. Face Validity: A Critical but Ignored Component of Scale Construction in Psychological Assessment. *Eur. J. Psychol. Assess. Off. Organ Eur. Assoc. Psychol. Assess.* **2023**, *39*, 153–156. [CrossRef]
46. Fowler, S.; Sitnikova, E. Toward a framework for assessing the cyber-worthiness of complex mission critical systems. In Proceedings of the 2019 Military Communications and Information Systems Conference (MilCIS), Canberra, Australia, 12–14 November 2019.
47. Fowler, S.; Joiner, K.; Sitnikova, E. Assessing cyber-worthiness of complex system capabilities using MBSE: A new rigorous engineering methodology. *IEEE Syst. J.* 2022. *submitted*. Available online: https://www.techrxiv.org/users/680765/articles/677291-assessing-cyber-worthiness-of-complex-system-capabilities-using-mbse-a-new-rigorous-engineering-methodology (accessed on 25 April 2024).
48. Cyber Evaluation and Management Toolkit (CEMT). Available online: https://github.com/stuartfowler/CEMT (accessed on 25 April 2024).
49. Fowler, S. Cyberworthiness Evaluation and Management Toolkit (CEMT): A model-based approach to cyberworthiness assessments. In Proceedings of the Systems Engineering Test & Evaluation (SETE) Conference 2022, Canberra, Australia, 12–14 September 2022.
50. National Institute of Standards and Technology (NIST) Computer Security Resource Center (CSRC), NIST Special Publication 800-160 Rev. 2: Developing Cyber-Resilient Systems: A Systems Security Engineering Approach. Available online: https://csrc.nist.gov/pubs/sp/800/160/v2/r1/final (accessed on 25 April 2024).
51. National Institute of Standards and Technology (NIST), CSF 2.0: Cybersecurity Framework. Available online: https://www.nist.gov/cyberframework (accessed on 25 April 2024).
52. Madni, A.; Purohit, S. Economic analysis of model-based systems engineering. *Systems* **2019**, *7*, 12. [CrossRef]
53. Bussemaker, J.; Boggero, L.; Nagel, B. The agile 4.0 project: MBSE to support cyber-physical collaborative aircraft development. *INCOSE Int. Symp.* **2023**, *33*, 163–182. [CrossRef]
54. Amoroso, E.G. *Fundamentals of Computer Security Technology*; Pearson College Div: Englewood Cliffs, NJ, USA, 1994.
55. INCOSE. *Systems Engineering Vision 2020*; International Council on Systems Engineering: Seattle, WA, USA, 2007.
56. Madni, A.M.; Sievers, M. Model-based systems engineering: Motivation, current status, and research opportunities. *Syst. Eng.* **2018**, *21*, 172–190. [CrossRef]
57. Huang, J.; Gheorghe, A.; Handley, H.; Pazos, P.; Pinto, A.; Kovacic, S.; Collins, A.; Keating, C.; Sousa-Poza, A.; Rabadi, G.; et al. Towards digital engineering—The advent of digital systems engineering. *Int. J. Syst. Syst. Eng.* **2020**, *10*, 234–261. [CrossRef]
58. Chelouati, M.; Boussif, A.; Beugin, J.; El Koursi, E.-M. Graphical safety assurance case using goal structuring notation (gsn)–challenges, opportunities and a framework for autonomous trains. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108–933. [CrossRef]
59. Sujan, M.; Spurgeon, P.; Cooke, M.; Weale, A.; Debenham, P.; Cross, S. The development of safety cases for healthcare services: Practical experiences, opportunities and challenges. *Reliab. Eng. Syst. Saf.* **2015**, *140*, 200–207. [CrossRef]

60. Nguyen, P.H.; Ali, S.; Yue, T. Model-based security engineering for cyber-physical systems: A systematic mapping study. *Inf. Softw. Technol.* **2017**, *83*, 116–135. [CrossRef]
61. Geismann, J.; Bodden, E. A systematic literature review of model-driven security engineering for cyber–physical systems. *J. Syst. Softw.* **2020**, *169*, 110697. [CrossRef]
62. Carter, B.; Adams, S.; Bakirtzis, G.; Sherburne, T.; Beling, P.; Horowitz, B. A preliminary design-phase security methodology for cyber–physical systems. *Systems* **2019**, *7*, 21. [CrossRef]
63. Larsen, M.H.; Muller, G.; Kokkula, S. A Conceptual Model-Based Systems Engineering Method for Creating Secure Cyber-Physical Systems. *INCOSE Int. Symp.* **2022**, *32*, 202–213. [CrossRef]
64. Japs, S.; Anacker, H.; Dumitrescu, R. SAVE: Security & safety by model-based systems engineering on the example of automotive industry. In Proceedings of the 31st CIRP Design Conference, Online, 19–21 May 2021.
65. Navas, J.; Voirin, J.; Paul, S.; Bonnet, S. Towards a model-based approach to systems and cybersecurity: Co-engineering in a product line context. *Insight (Int. Counc. Syst. Eng.)* **2020**, *23*, 39–43. [CrossRef]
66. Geismann, J.; Gerking, C.; Bodden, E. Towards ensuring security by design in cyber-physical systems engineering processes. In Proceedings of the International Conference on the Software and Systems Process, Gothenburg, Sweden, 26–27 May 2018.
67. Mažeika, D.; Butleris, R. MBSEsec: Model-based systems engineering method for creating secure systems. *Appl. Sci.* **2020**, *10*, 2574. [CrossRef]
68. Object Management Group. UAF: Unified Architecture Framework. 2022. Available online: https://www.omg.org/spec/UAF. (accessed on 15 June 2024).
69. Jurjens, J. *Secure Systems Development with UML*; Springer: Berlin/Heidelberg, Germany, 2005.
70. Apvrille, L.; Roudier, Y. Towards the model-driven engineering of secure yet safe embedded systems. *Int. Workshop Graph. Models Secur.* **2014**, *148*, 15–30. [CrossRef]

CHAPTER 6 IMPLEMENTATION TRIALS

6.1 EMBEDDED CHAPTER – COMPUTERS AND SECURITY

This chapter has been submitted for publication as

Fowler, S.; Joiner, K.; Ma, S. Assessing Cyberworthiness of Complex System Capabilities using the Cyber Evaluation and Management Toolkit (CEMT), *preprint* [submitted]

Assessing Cyberworthiness of Complex System Capabilities using the Cyber Evaluation and Management Toolkit (CEMT)

Stuart Fowler[a], Keith Joiner[b] and Siqi Ma[a]

[a] School of Systems and Computing, University of New South Wales, Canberra, Australia
[b] Corresponding Author, email k.joiner@unsw.edu.au
Capability Systems Centre, University of New South Wales, Canberra, Australia

Abstract—Cyberworthiness, as it is termed in Australian Defence, is a necessary feature of modern complex systems that are required to operate in a hostile cyber environment. To evaluate the cyberworthiness of complex systems or system-of-systems, the researchers developed a Cyber Evaluation and Management Toolkit (CEMT) which leverages model-based systems engineering techniques to examine a cyber-physical system's vulnerability to, and risk of, cyber-attacks. This assessment method addresses the cyber attack surface and threat kill chains, including supply chains and supporting infrastructure. Our previous research developing the CEMT is summarised and the research results of an implementation trial on a naval capability are presented. CEMT is novel to other similar assessment approaches articulated in the literature research by being threat-led and using visualisations tailored to inform decision-makers without specific cyber expertise. The results of the research trial are promising and provide strong justification for the continued application of the CEMT when evaluating and assessing the cyberworthiness of complex cyber-physical systems.

Keywords—Cyberworthiness, Cybersecurity, Model-Based Systems Engineering, Complex Cyber-Physical Systems, Case Study, Operational Evaluation

1 INTRODUCTION

Cyberworthiness is a term coined within the Australia Department of Defence (DoD) [1] and is "*a measure of a capability's suitability to operate in its intended environment*," [2] where that intended environment includes a contested cyber domain. The intent of cyberworthiness is broadly aligned with cyber resiliency [3] and cyber survivability [4], which are more commonly used within the National Institute of Standards and Technology (NIST) family of cybersecurity standards. However, the deliberate decision to invoke the concepts of '*worthiness*' with the cyberworthiness term leverages the history of existing regulatory systems such as airworthiness [5, 6], seaworthiness [7] and landworthiness [8], which focus on a capability's suitability to safely and effectively operate in the traditional air, sea or land domains.

A claim of worthiness requires that the approach to decision making, not just the decision itself, is sound. As outlined in the *Principles for Governance of Seaworthiness* from the Defence Seaworthiness Management System Manual [9]:

> *"2.10 The seaworthiness governance principles require that seaworthiness decisions are made:*
> *a. mindfully – decisions are more effective and less likely to have unintended consequences when they are made with a thorough understanding of the context, the required outcome, the options available, and their implications now and in the future*
> *b. collaboratively – obtaining input from all stakeholders and engaging in joint problem-solving results in better decisions (bearing in mind that collaboration does not necessarily require consensus)*
> *c. accountably – decisions only become effective when people take accountability for making them happen*
> *d. transparently – decisions are more effective when everyone understands what has been decided and why." [9] (p.33)*

Applying these same principles to the cyber domain, a claim of cyberworthiness therefore requires that decision-makers have a clear understanding of the risk that they are accepting, and that the evidence justifying those decisions is clearly identifiable and traceable to the decision. Such provisions help ensure vicarious liability would be extended to certifiers in the event new cyber-threats result in safety consequences.

Cyberworthiness is a departure from the traditional approach to assessing cybersecurity risk within the Australian DoD. This traditional approach follows the six-step process of the Information Security Manual (ISM), which selects and implements cybersecurity controls from the ISM, assesses their effectiveness and authorises a residual risk based on the effective implementation of those controls [2]. While this isn't strictly a compliance-based approach, the risk assessment is often based on the level to which compliance is achieved against the ISM control baseline – an audit against the control baseline is undertaken, and any controls that are not effectively implemented contribute to the

qualitative residual risk level that is assessed. This audit-focused approach is not readily conducive to the cyberworthiness principles outlined above, necessitating a new approach to assess cyberworthiness risk.

This demand signal for a new way of assessing risk isn't unique to Australia. The US Department of Navy (DON) has also identified that a new approach to cybersecurity is required for their systems:

> *"The DON needs a new approach to cybersecurity that goes beyond compliance because our over-reliance on compliance has resulted in insecure systems, which jeopardize the missions these systems support. Instead of a compliance mindset, the DON will shift to Cyber Ready, where the right to operate is earned and managed every day. The DON will make this transition by adhering to an operationally relevant, threat-informed process that affordably reduces risk and produces capabilities that remain secure after they have been delivered at speed." [10] (p. 7)*

A modern approach that is threat-informed and provides more effective risk assessments due to its focus on the operationally relevant risks is clearly desired.

The efficiency of any new approach is also paramount. The recent audit of the Australian DoD by the Australian National Audit Office (ANAO) identified that "*the volume of Defence ICT systems outweighs available resources for assessment and authorisation activities*" [2] (p. 82). This was particularly bad for Navy systems, where less than 1 per cent of Navy-owned ICT and embedded systems (i.e. operational technology) of maritime material and support systems had been assessed and authorised [2] (p. 56). Given the scope of the problem, an inefficient methodology is not going to be able to resolve the extant issues.

This efficiency is likely to only be reached through reusability and consistency of the assessment and assurance activities between similar systems. A new approach that requires a bespoke threat model to be generated for each system is unlikely to reach the efficiency required to adequately deal with the vast number of systems that require assessment. It is therefore key to ensure that the work performed using any new approach addresses the need to leverage previous assessments in future assessments while still remaining contextually relevant for each system being assessed.

Our earlier research [12] articulated how model-based security assessment approaches are ideally positioned to address this demand for a more efficient and effective methodology for the assessment of cyberworthiness. The developmental research defined five key attributes that contribute to a comprehensive cyberworthiness evaluation using model-based security assessments, namely:

1. Extended Model-based Taxonomy – an extension of an open model-based systems engineering language, such as UML or SysML, to facilitate a model-based approach.
2. Threat Focused – the threats to the system, rather than a best-practice control baseline or asset hierarchy, is used as the focal point of the assessment.
3. Detailed Adversary Modelling – the actions of the adversary are modelled in detail, facilitating a precise discussion and review of any threat analysis.
4. Visualisation and Simulation of Threat – detailed adversary modelling is expressed in simplified graphs, such as *attack trees*, and branches of those graphs can be simulated quantitatively.
5. Explicit Traceability to Threats – derived security controls are directly traceable to adversary actions, facilitating discussion and review of the importance of each control in terms of the malicious action it mitigates

2 LITERATURE

A summary of the degree to which the existing research literature of model-based security assessment approaches address these five key attributes is shown in Table 1. This table is updated from our earlier work to cover recent advances in the field, specifically the extension of Larsen et al. [16] to include explicit traceability between security controls and threats, as well as the translation of the popular *System-Theoretic Process Analysis for Security* (STPA-Sec) approach [24] into a model-based security assessment [25]. Following these updates, there remains no other published approach that addresses all five of these key attributes.

Table 1. Feature Presence Comparison of model-based security assessments along five key attributes

	Model-based Security Assessment Approach	Extended Model-based Taxonomy	Threat Focused	Detailed Adversary Modelling	Explicit Traceability to Threats	Visualisation & Simulation of Threats
1	*CSRM* [13]	YES	NO	NO	NO	NO
2	Larsen et al. [15] [16]	YES	YES	NO	YES	NO
3	*SAVE* [17]	YES	YES	NO	NO	NO
4	Navas et al. [18]	YES	YES	NO	NO	NO
5	Geissman et al. [19]	YES	YES	NO	NO	NO
6	*MBSESec* [20]	YES	YES	YES	NO	NO
7	*UAF* [21]	YES	NO	NO	NO	NO
8	*UMLSec* [22]	YES	NO	NO	NO	NO
9	*SysML*-Sec [23]	YES	NO	NO	NO	NO
10	STPA-Sec in MBSE [25]	YES	NO	NO	NO	NO

Most of these model-based security assessment approaches are published in the literature using author-defined examples, with the peer review process serving as the primary validation of the approach. Several of these approaches, specifically *SAVE* [17], *MBSESec* [20] and the recent publication by Larsen et al. [16] use face-validity trials based on participant surveys following case study implementations to validate the efficacy of those approaches. While STPA-Sec itself has had a significant body of research validating its use in the field [26], these trials use the underlying STPA-Sec technique, rather than the translation of that technique into a model-based approach [25].

This absence of published work-based studies into the implementation of these model-based security assessment approaches represents a gap in the overall literature. Applied field studies will help us understand and validate the efficiency and effectiveness of the approaches under real-world conditions and provide a better foundation for evaluating the real-world user satisfaction with such approaches.

To address this gap, our research questions for this investigation were:

1. How effective is our developed model-based security assessment approach at assessing the cyberworthiness of a real-world cyber-physical system?
2. How efficient is our developed model-based security assessment approach at assessing the cyberworthiness of a real-world cyber-physical system?
3. How well does our developed model-based security assessment approach inform decision-makers of the key risks associated with the real-world cyber-physical system?

3 METHODOLOGY

Building on our early aims [11] and over six years of research on best-practice in this field, we developed an approach for leveraging model-based systems engineering techniques to perform threat-informed risk assessments called the Cyber Evaluation and Management Toolkit (CEMT) [12]. The aim was to meet all five of the identified key attributes that contribute to a comprehensive cyberworthiness evaluation using model-based security assessments. As detailed in our earlier work, and on the CEMT website [13], the toolkit facilitates the following steps:

1. Identification of *misuse cases*, which are the high-level threats to the system under assessment.
2. Articulation of '*mal-activity*' diagrams, which detail the steps an adversary would need to take to compromise the system under assessment.
3. Derivation of a security control baseline, based on the mitigations that could be applied at each node of the *mal-activity diagram* to reduce the risk of compromise.
4. Production of '*attack trees*,' which summarise the adversary attack surface and identification of threat vectors into the system under assessment.
5. Selection of critical threat vectors, using qualitative assessment techniques and stakeholder engagement to prioritise the branches of the *attack tree*.

6. Quantitative assessment of key risks, by assigning initial probabilities to each critical threat vector and assigning control effectiveness values to each of the controls sets at each step in the risk's kill chain informed appropriately by threat-intelligence.
7. Mapping quantitative simulation results and translating the results of the numerical simulation into qualitative likelihood bands which are combined with a consequence rating to determine inherent risk.
8. Proposing additional mitigations against each of the highest risks sensible to their context and then re-simulating to estimate the residual risk following implementation of those additional mitigations.

Following the development of the CEMT, a face-validity investigation was conducted to assess the suitability of the toolkit for evaluating the cyberworthiness of complex systems [12]. This investigation involved providing presentations and demonstrations of the toolkit to subject matter experts and surveying them on their initial impressions. The results of this survey were promising and provided strong justification for proceeding to field research of the toolkit.

Since this earlier work, the CEMT has been applied to a naval platform to support the assessment and authorisation of that capability [2]. This research trial was performed between March 2024 and November 2024, to assess the cyber risk, current controls and mitigation strategies of a capability. Stakeholders in this implementation trial were given the same survey used in the face-validity trials to evaluate the effectiveness, efficiency and user satisfaction with the CEMT as a tool for assessing the cyber-worthiness of a complex cyber-physical mission system [12].

The survey asked questions about the CEMT methodology itself, the quality of the outputs from the process, the impact those outputs can have on the decision-making process and the level to which it supports continuous assurance in an evergreen environment. Respondents were also asked to provide their overall impression of the approach taken during the implementation trial. The survey was distributed to all stakeholders in the trial, including the security assessors performing the threat modelling, cybersecurity experts and system designers that were involved in the peer review of the assessment, and decision-makers who were ultimately responsible for acting on the results of the assessment.

The field research trial ran for nine months and included the entire CEMT process, from identifying *misuse cases* and articulating *mal-activity diagrams* to generating *attack trees* and simulating risk assessments. This research trial was the first time the CEMT was applied on a large, complex system and because of this several improvements and enhancements were made to the CEMT during the implementation trial to allow the tool to scale more effectively. These represent additional extensions from the previously published CEMT capabilities, but they were part of the CEMT, namely its adaptability, that was evaluated through the survey process. Those upgrades, and the implementation challenges that drove them, are articulated in the following section.

4 Implementation Challenges

4.1 Scaling Difficulties

The first challenge that was faced during the implementation of the CEMT on a complex, real-world system was the need to streamline the development of the threat model through both internal and external reuse. The previously published CEMT capability required a threat model to be developed that articulated the steps an attacker would need to take to compromise the system, and then those steps to be linked to both the assets that are impacted by that threat step, and the controls that could mitigate that step.

This was acceptable on a small sample system, but when applied to a large, complex system the time taken to link each asset to each threat step individually was cumbersome and time consuming, of the order of hundreds of thousands of links. To address this, an asset group stereotype was added to the CEMT, which can be used to group assets based on modeler-defined attributes of the assets. As an example, an asset group of '*Has USB Ports*' can be created and linked to threat model steps that are performed against assets with USB ports. Where multiple asset groups are linked to single threat steps, the asset needs to be in all groups for it to be considered as affected by that threat step. As assets are added to the model for a particular system, they can be allocated into asset groups based on their attributes, and this creates an indirect relationship to the threat steps via the asset groups. Figure 1 and Figure 2 outline the difference in the legacy approach which directly linked assets (Figure 1) and the updated approach which uses asset groups (Figure 2).

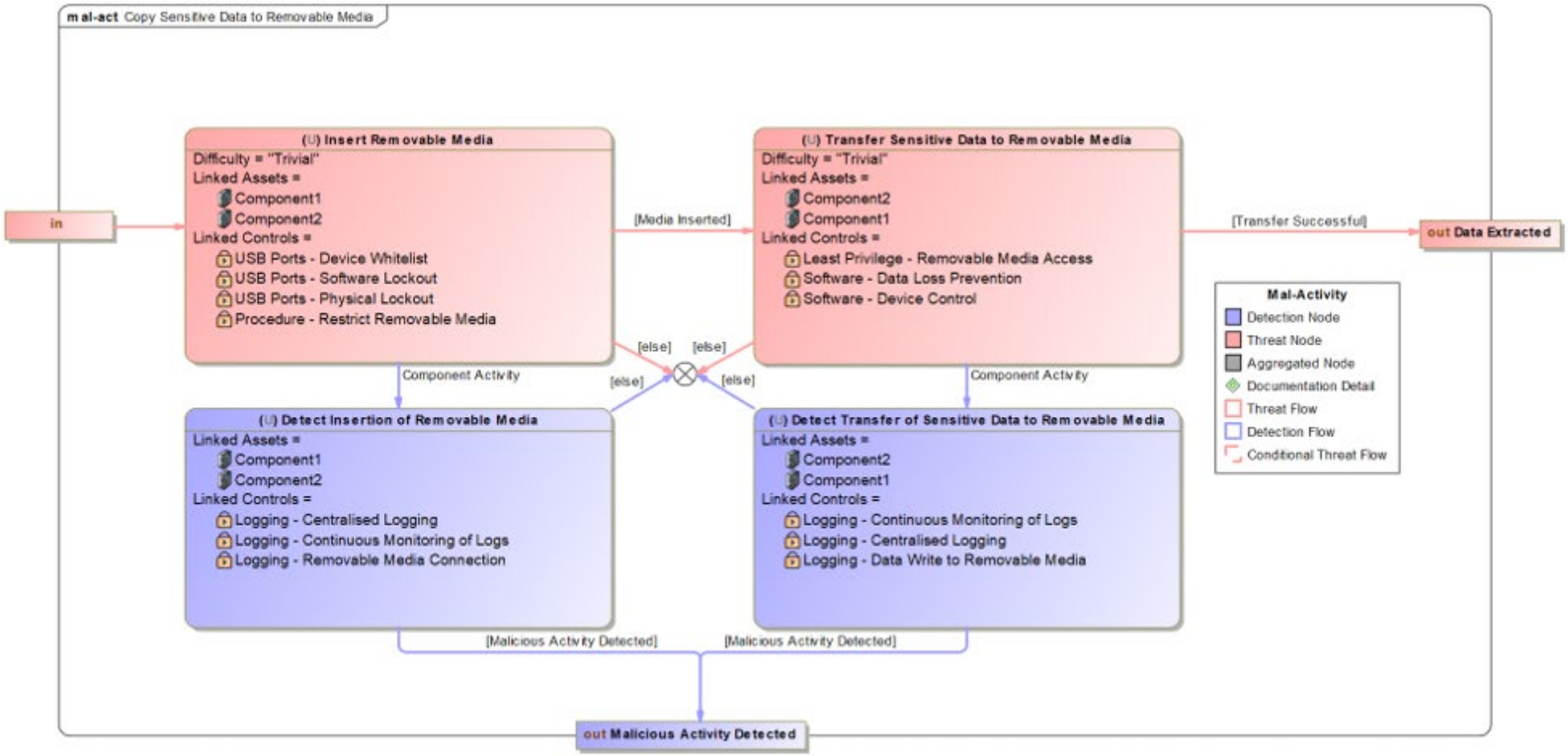


Figure 1. *CEMT Mal-Activity Diagram* showing assets (*Component1, Component2*) directly linked to the threat steps

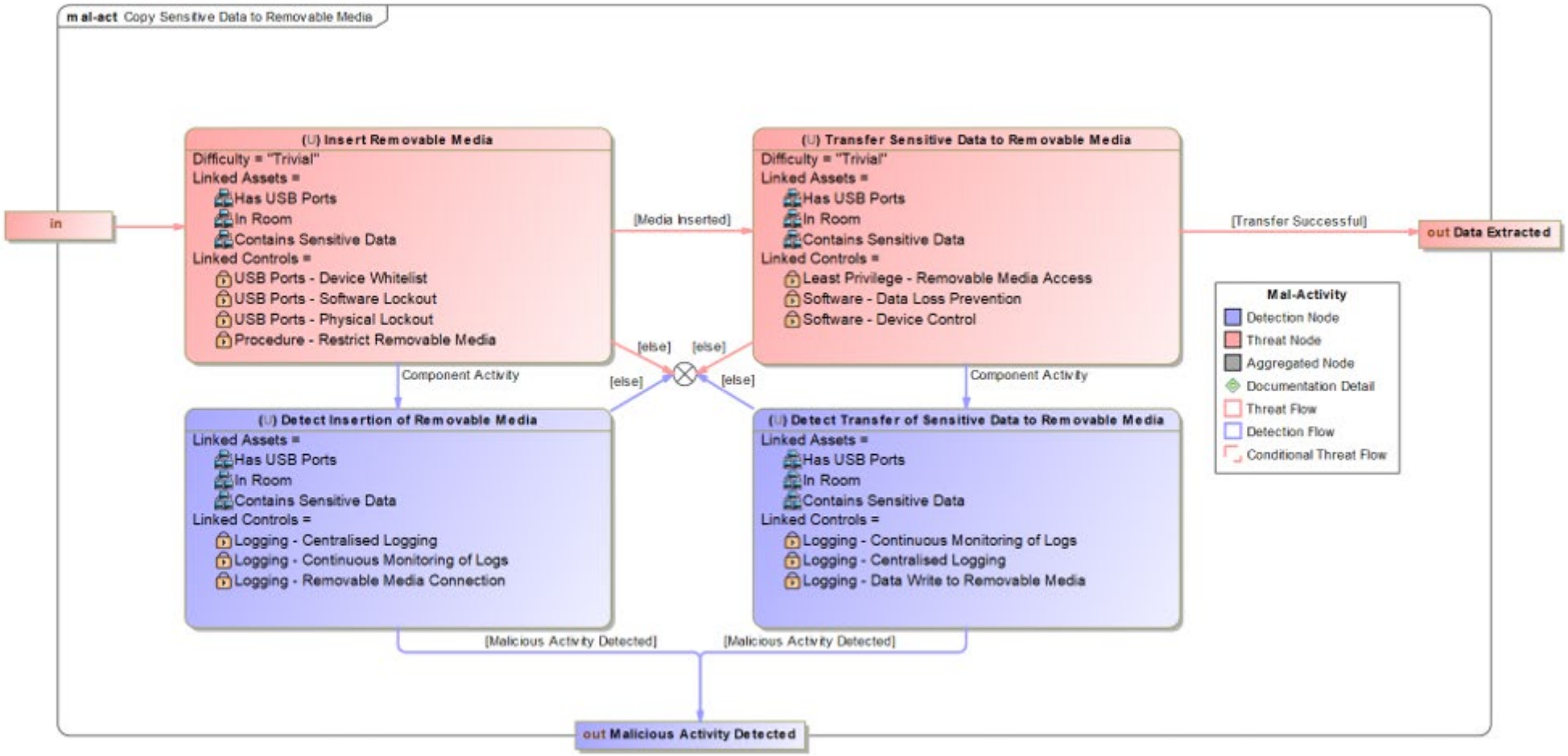


Figure 2. *CEMT Mal-Activity Diagram* showing asset groups (*Has USB Ports, Has User Interface, Contains Sensitive Data*)

The additional benefit of having the threat model linked to conceptual asset groups rather than being directly linked to system-specific assets is that the *misuse cases* and *mal-activity diagrams* are now portable between different assessments on different systems. This new feature means that reusable, generic threat sets can be developed that include both the control sets and the asset groups as a centralised library within an organisation or industry. To apply these threats to new systems, the generic threat set can be imported into a model and the assets of that specific system linked into the relevant asset groups that come with the generic threat set. This allows for *misuse cases* to be centrally managed and allocated down to systems, greatly increasing the reusability of CEMT-developed threat models.

As an example, a generic threat set for network penetration attacks can be developed under a single *misuse case* which contains *mal-activity diagrams* that align to the tactics and techniques of the *MITRE ATT&CK Framework* [27]. The threat steps in these *mal-activity diagrams* can then be linked to system-agnostic asset groups and security controls drawn from both the *MITRE ATT&CK Framework*, and other control frameworks such as the ISM [28] and NIST 800-53 [29]. This creates a portable generic threat set for a Network Penetration *misuse case*, that contains the traceability between an *attack tree* and best-practice mitigations, which can be imported into any system to which this threat set is applicable. The asset hierarchy for that new system can be modelled and those assets can be allocated to the asset groups from the generic threat set based on their attributes, such as whether they are network connected, or have USB ports. This approach will indirectly link the assets into the generic threat set based on the direct asset group relationships, and a threat-informed control baseline will be derived for network penetration attacks against the system being assessed. Furthermore this approach allows for a flexible first-principles approach to be used to develop bespoke or advanced threat sets on a system-by-system basis, while allowing for those common or generic threat sets to be easily reused and reapplied across similar systems.

The system assessed during the implementation trial produced a threat model with over 200,000 different branches on the *attack tree* and hundreds of individual assets, which overwhelmed the ability to manually and exhaustively perform a qualitative assessment across the entire *attack tree* to determine which threat vectors should be quantitatively assessed and which mitigations were most critical. Consequently, the CEMT was updated with the ability to provide an automated assessment of the relative importance of mitigation and detection controls using a combination of the prevalence of the control on the *attack tree* and the criticality of the asset on which the control was implemented. The algorithm used to create this relative importance (I_x) is provided in equation (1):

$$I_x = \left(\frac{n_x}{MAX(n_1, n_2, \dots, n_y)} \right) * 100 \quad (1)$$

Where n_x represents the number of unique *attack tree* branches the security control is mitigating in the threat model and y represents the number of unique security controls in the model. Therefore I_x represents the relative importance of the control based on the breadth of the applicability of the control across the attack surface, normalised to the control with the highest breadth of applicability.

4.2 Risk Categorisation and Modularisation

The large *attack tree* also led to many threat vectors being identified for quantitative assessment. In the previous versions of the CEMT, these risks were simply categorised by the *misuse cases* from which they start, but with a larger number of risks being identified, a more hierarchical categorisation method was required. The CEMT was updated to align with the '*Mission, Attack and Variant structure*' used by the US DoD in their Cyber Table-Top guide [30]. *Missions* represent the goal of the adversary attempting to compromise the system, and is usually expressed as the denial or degradation of critical functionality. *Attacks* are subordinate to the *Missions* and represent attacks on specific components or subsystems which contribute to the function the mission is trying to impact. Finally, *Variants* are the different ways in which the *Attack* could be carried out and have a direct relationship with a branch of the *attack tree*. This creates a tree hierarchy that effectively categorises each of the individual risks into groups of similar threat vectors.

The individual risks then become slight variations of each other, and it was noticed during the implementation trial that this led to a significant amount of duplication in the quantitative risk assessments. To resolve this issue, the CEMT was updated to make better use of the object-oriented nature of the model-based systems engineering tool by introducing '*risk modules*,' which are segments of the *attack tree* which can be combined to make the overall parametric risk diagram. This change moves the control effectiveness determination into a reusable module that can be imported into each parametric risk simulation, reducing the duplication and eliminating the risk of misalignment. This refactoring of the approach to risk calculation necessitated a change in the way the detection probability was calculated in the parametric simulations. The previous approach was to calculate the probability of detection at each node individually

and then do probability addition once all detection probabilities had been calculated, as described in our previous research [12]. In contrast, the new modularised risk approach necessitated an alternative algorithm which calculates the running total of the detection probability up to that point in the threat path so that it could be carried through the chain of risk modules. This algorithm is explained in equation (2):

$$P_{xD} = P_{(x-1)D} + p_{xD} \tag{2}$$

Where P_{xD} is the cumulative detection probability after x steps in the threat vector, and p_{xD} is the probability that an attack that has not previously been detected is detected at the xth step in the threat vector. P_{0_D}, the initial value for this cumulative detection probability, is defined as zero per cent.

Equation (3) shows that p_{xD} is calculated by multiplying the probability that an attack had not been detected prior to this node $P_{(x-1)ND}$, the probability the attack reached this node P_{xT}, the percentage of attempted attacks that would be detected by the implemented controls for this node E_{xD} and the percentage of potential detections that would not be avoided by the attacker A_{xD}:

$$p_{xD} = P_{(x-1)ND} \times P_{xT} \times E_{xD} \times A_{xD} \tag{3}$$

E_{xD} and A_{xD} are user-defined variables set during the modelling process using expert judgement or historical data. P_{xT} is calculated using the existing algorithm for calculating residual probability at each node, as described in our previous work [12]. The probability that an attack has not been detected after node x is calculated using equation (4) below:

$$P_{xND} = P_{(x-1)ND} \times P_{xT} \times [1 - (E_{xD} \times A_{xD})] \tag{4}$$

$P_{0_{ND}}$, the initial value for this progressive probability that an attack has not been detected is defined as 100%. This new algorithm allows the detection probability calculation to be chained together, which simplifies the diagrams for connecting the new modularised parametric risk diagrams together.

The application of the CEMT on a real-world implementation also identified that while the CEMT was able to be updated to address changes to the threat environment or to the system design over time, there was also a need to simultaneously assess multiple threat environments and multiple design variants.

4.3 THREAT ENVIRONMENT VARIATIONS

The implementation trial was performed on a naval capability, and the threat environment for a warship is markedly different when it is in harbour and when it is at sea. As a contrived example, the likelihood of a malicious actor walking onto the ship as the start of an attack is very different when the ship is alongside or in maintenance than when it is at sea; however, the threat model needs to be able to address all states and modes of the system simultaneously to make a comprehensive assessment. To address such use cases, the concept of an '*Operation*' was introduced to the CEMT, which allows risks to be re-instantiated and re-simulated with different initial probabilities. These re-simulations are saved alongside the baseline assessment and can be used to articulate how the risk changes when the system performs different actions or is operating in a particular state or mode.

The ability to handle design variants was also identified as a desirable addition to the CEMT during the implementation trial. Specifically on naval capabilities, there are often multiple ships in a single class of ship, but each of those individual ships has slightly different design baselines due to the time-phased deployment patterns of updates, uplifts and upgrades to the ships within the class. While each design variant could have been assessed as a separate CEMT model, the CEMT was updated during the research field trial to include support for '*Platforms*,' which are instances of the system under test that can be associated with a subset of the risks in the threat model. This addition allows a modeler to create an asset hierarchy that is a superset of each design variant, build risks based on a valid collection of assets and then allocate only the relevant risks to each platform. This approach allows for maximum reuse of the threat model, while still incorporating the minor differences between ships of the same class.

5 FINDINGS

As described in Section 3, following the completion of the research field trial a survey was distributed to all participating stakeholders to solicit their feedback on the use of the CEMT in the implementation trial. The respondents

were asked to provide answers to 41 specific questions on a five-point Likert scale. Questions 1 to 11 focused on the primary attributes of the CEMT methodology, questions 12 to 23 focused on the output produced by the CEMT during the implementation trial, questions 24 to 31 focused on the ability for the CEMT and its outputs to support decision-making, questions 32 to 36 focused on the ability for a CEMT-based assessment to support the evergreen continuous assurance of the system, and questions 37 to 41 focused on the overall impressions of the CEMT.

The demographics of responses were primarily technical specialists with a background in the development of cyber-physical systems. A total of 32 responses to the survey were received (n = 32), 18 of which said they were from *Defence*, nine from *Industry* and five from *Government*. The majority of responses stated that their relationship to the development of cyber-physical systems was as a *Technical Specialist* (20), followed by *Governance* (7), *Management* (4) and *Other* (1). 25 respondents identified themselves as cybersecurity experts, with experience ranging from one year of experience to 25 years of experience, with a mean of 6.76 years. The primary expertise of the participants included *IT / Computing* (14), *Engineering* (13), *Operations* (3) and *Management* (2). There was participation from all key roles in the CEMT process, including *Assessor / Modeller* (6), *System Designer / Technical Lead* (4), *Peer Reviewer - Non-Cyber Expertise* (5), *Peer Review - Cyber Expertise* (15) and *Decision Maker* (2).

Figure 3 provides the results of the overarching opinion of the respondents to the CEMT process as a whole. 62.5% of respondents provided an overall rating of *Excellent* or *Very Good* to the CEMT as a cyberworthiness assessment methodology, and just over 65% indicated that they *Strongly Agree* or *Agree* that the CEMT is an appropriate technique for assessing the cyberworthiness of systems they work on. Approximately 30% of respondents provided neutral response to these questions, which may indicate an unwillingness to commit to a strong position on these overarching conclusions after participating in a single research trial.

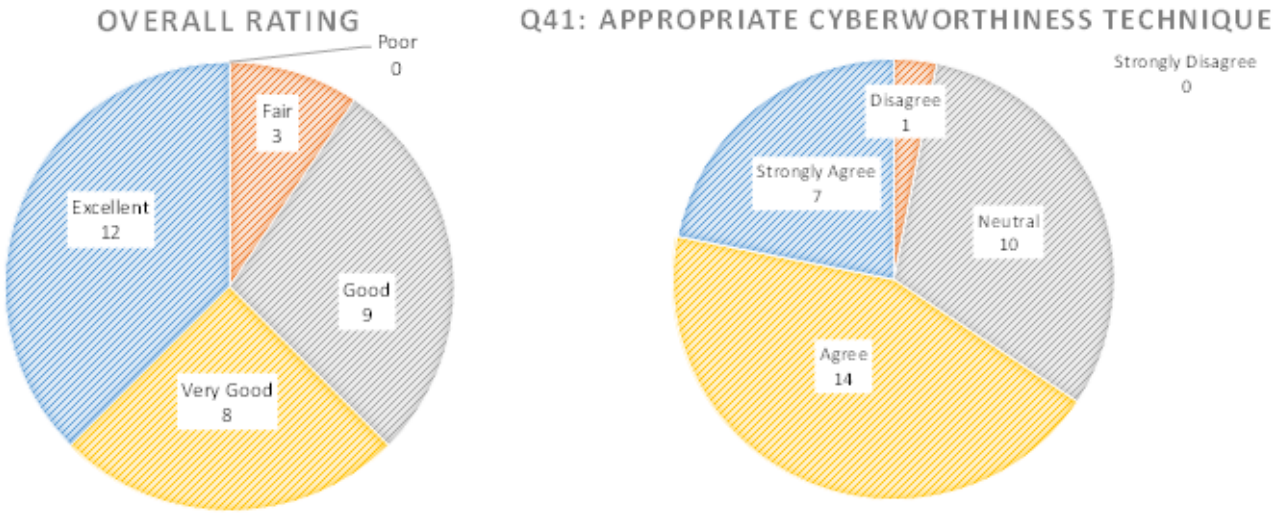


Figure 3. Appropriateness and Overall Rating

The detailed results of the survey are shown in Figure 4 and Figure 5.

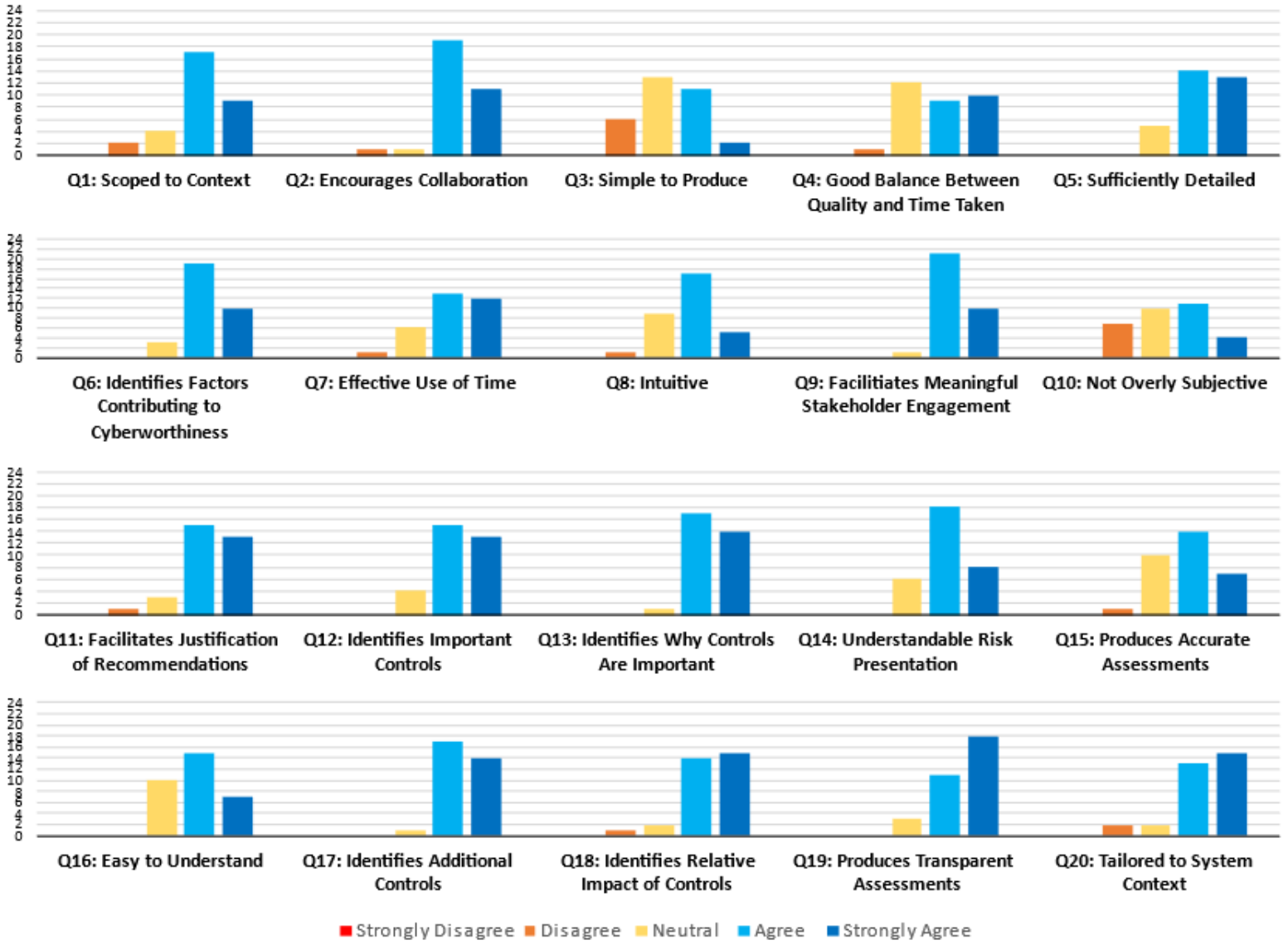


Figure 4. Results Summary (Q1 – Q20)

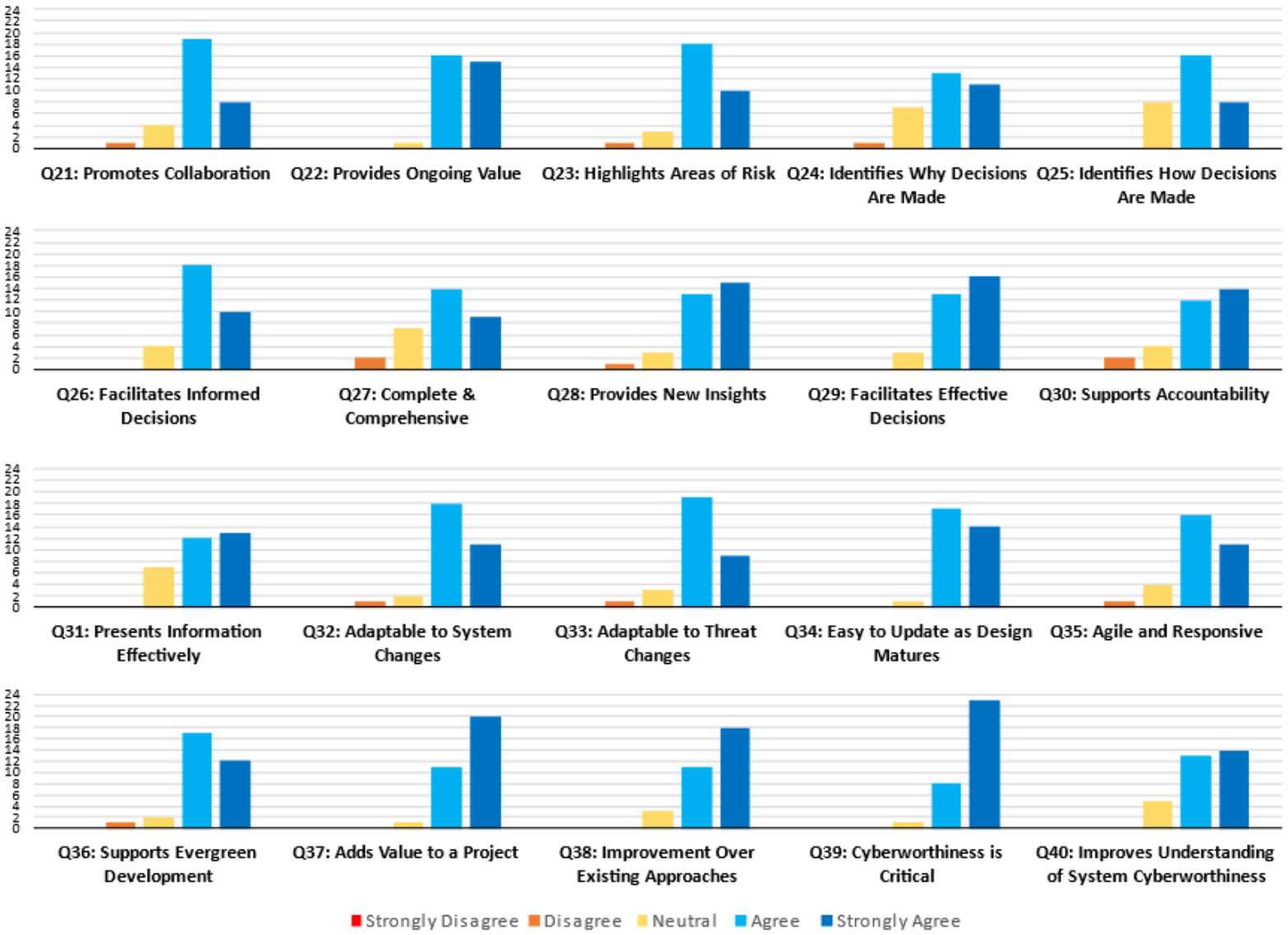


Figure 5. Results Summary (Q21 - Q40)

The detailed results show a generally favourable evaluation of the CEMT, with positive responses outweighing the negative responses for every question in the survey. There are some notable outliers, where there were more than two respondents who provided a disagreeable response to the question. These include:

- **Q3**: "*Cyberworthiness assessments are simple to produce using the CEMT*" and
- **Q10**: "*The CEMT is not overly dependent on the subjective opinion of subject matter experts.*"

The demographic breakdown for the responses to Q3 is worth examining more closely, as not all respondents were involved in the production of the cyberworthiness assessment using the CEMT and so their direct experience on which to base a judgement as to the simplicity of production is varied. As noted earlier, of the 32 respondents only six identified their role in the assessment as a '*Assessor / Modeller*' and this was the only role that actively produced the assessment; the other roles in the assessment were either reviewers, decision-makers or system designers. These six respondents who were directly involved in the production of the CEMT assessment responded to Q3 with 1 *Neutral*, 3 *Agrees* and 2 *Strongly Agrees*. This result may suggest the CEMT assessment looked more difficult to produce than it was to produce, or it could simply indicate that those respondents had more time to become familiar with the CEMT.

The results for Q10 likely reflect that the simulations and quantitative assessment of the CEMT assessment is highly sensitive to the input values for initial probability and control effectiveness. In the absence of strong historical data, these assessments will remain subjective decisions from subject matter experts and so this finding is not an unexpected one – our hypothesis was that expert judgement would not be eliminated using the CEMT approach, but that we would be able to achieve a reduction in opaque expert judgement. The favourable results for Q19 *Produces Transparent Assessments* — as well as Q24 *Identifies Why Decisions Are Made* and Q25 *Identifies How Decisions Are Made* — indicate that this reduction in opaque expert judgement may have been achieved.

The respondents were also asked how likely they were to recommend CEMT on a scale of 0-10, with 0 the least likely and 10 the most likely. The mean response to this question was 8.6, with 3 respondents falling into the *Detractors* band (6 or below), 10 falling into the *Passive* band (7 or 8) and 19 falling into the *Promoters* band (9 or 10).

6 DISCUSSION

6.1 EFFICIENCY

The results of the research field trial demonstrate that the CEMT can provide an efficient method for assessing the cyberworthiness of complex cyber-physical systems. While there was some disagreement on Q3 *Simple to Produce*, which is a key component of efficiency, there were more positive responses to Q4 *Good Balance Between Quality and Time Taken* and Q7 *Effective Use of Time*. This result indicates that the research trial participants agreed that while the completion of a cyber-worthiness assessment using the CEMT was not necessarily easy or simple, the time and effort that were spent on the assessment were worthwhile, which is a key component of an efficient methodology.

Efficiency isn't only confined to the initial development of the CEMT threat model. Q1 *Scoped to Context* and Q20 *Tailored to Context* received broad agreement, which indicates that nugatory effort was not spent assessing irrelevant security threats or mitigations. These questions are intended to explore the potential benefits of the threat-focused approach inherent in the CEMT.

Existing security approaches for the cyber assessment of complex systems in the Australian DoD, as mentioned in Section 1, focus on the application of best-practice cybersecurity controls from the Information Security Manual. The security controls in the ISM are intended to apply to all ICT systems across the Australian Government, which means they primarily focus on appropriate controls for enterprise ICT systems. However, the best practice for securing enterprise ICT systems is not necessarily best-practice for securing a naval platform, and significant time can be spent implementing, auditing and assessing the effectiveness of controls that aren't necessarily mitigating security risk to a cyber-physical system given the different context in which those systems operate. The positive response to Q1 *Scoped to Context* and Q20 *Tailored to Context* highlights that the CEMT focuses on those threats and mitigations most relevant to the system under assessment, and in doing so, the efficiency of the overall assessment is improved.

Similarly, efficiency of the overall assessment can be influenced by the efficiency with which the initial assessment is updated and maintained throughout the life of the system. A complex cyber-physical system by definition is time-variant, either due to upgrades to the system design, changes to the operating context, or changes to the threat environment. Under the assumption that the residual risk needs to be understood following anyone of these changes, the inertia that needs to be overcome to update the cyberworthiness assessment directly contributes to the overall efficiency

of the assessment process. Survey respondents agreed that the CEMT process was Q32 *Adaptable to System Changes*, Q33 *Adaptable to Threat Changes* and Q34 *Easy to Update as Design Changes* — indicating that this efficiency through life would be achieved using the CEMT. It is important to note, however, that this implementation trial focused primarily on the application of the CEMT over the period of nine months, so there was limited ability for stakeholder participants to understand how well a CEMT approach would perform over the whole life of a capability.

6.2 EFFECTIVENESS

The findings from the research field trial also show that the respondents found the CEMT to be an effective process for assessing the cyberworthiness of a cyber-physical system. The responses to Q12 *Identifies Important Controls* and Q23 *Highlights Areas of Risk* highlight that the CEMT produces effective assessments that output two of the critical components of a cybersecurity risk assessment — the areas of risk and the controls that are mitigating those risks. These components were also described at an appropriate level of detail, and participants believed that the assessments were accurate, as seen in the responses to Q3 *Sufficiently Detailed* and Q15 *Produces Accurate Assessments*. These results indicate that the CEMT can effectively produce the primary output of a cybersecurity risk assessment in a detailed and accurate manner.

One of our key research questions was to understand how well the decision-makers are informed of the risk they are being asked to accept. The value of an accurate risk assessment is reduced if decision-makers are not able to interpret or understand the risk assessment in a way that allows them to appropriately balance the identified risk against other business and mission goals. The results for three questions showed broad agreement that the CEMT provides clear justifications for the cyberworthiness risk decisions that are supported by the CEMT: Q6 *Identifies Factors Contributing to Cyberworthiness*, Q24 *Identifies Why Decisions Are Made,* and Q25 *Identifies How Decisions Are Made*. Similarly, the positive sentiment of the respondents toward Q19 *Produces Transparent Assessments* highlights that the reasoning supporting those decisions are visible to both decision-makers and peer-reviewers. Ultimately, the combination of these factors led the respondents to believe that the CEMT approach was able to assist decision-makers to make both informed and effective decisions, as further shown in responses to Q26 *Facilitates Informed Decisions* and Q29 *Facilitates Effective Decisions*.

Unsurprisingly, the results also indicated that there is still some way to go with the effectiveness of the CEMT approach, particularly with respect to objectivity. The responses to Q10 *Not Overly Subjective* were mostly mixed, indicating that the research trial participants were less supportive of this statement than most of the other statements in the survey. This result suggests that there is additional work to be done to reduce this subjectivity further. The positive responses to Q21 *Promotes Collaboration*, combined with the transparency and visibility of the justifications for decisions outlined previously, suggest that the CEMT could support rigorous peer-review and group consensus to try and temper this subjectivity. However, the injection of objectivity into the assessments might require additional effort external to the CEMT to validate the input values for initial probability of a particular threat or the effectiveness of a particular control set through testing, monitoring and historical data analysis.

6.3 USER SATISFACTION

The results from the research field trial also showed strong positive results for user satisfaction with the CEMT process. The survey responses to Q14 *Understandable Risk Presentation* show that the risks were presented in a way that the participants could understand, and the responses to Q9 *Facilitates Meaningful Stakeholder Engagement* indicate that the research trial participants generally felt the process allowed them to take that understanding and meaningfully engage with the cyberworthiness assessment. When a broad stakeholder group is meaningfully engaged in the assessment process, they are more likely to trust the process and the outputs of that process.

This overall positive impression of the CEMT process was reflected in the participants' generally strong agreement with Q37 *Adds Value to a Project* and Q40 *Improves Understanding of System Cyberworthiness*. Ultimately, the participants also agreed with Q38 *Improvement Over Existing Approaches*, indicating the trial was successful in demonstrating that the CEMT can produce cyberworthiness assessments, one that assessment stakeholders feel is better than the current systems and processes.

7 CONCLUSION AND FUTURE DIRECTIONS

The CEMT was developed to evaluate whether a model-based systems engineering approach would be an efficient and effective method for assessing the cyberworthiness of complex cyber-physical systems. Face-validity trials were conducted where experts were given demonstrations of the CEMT and surveyed for their initial impressions [12]. These initial research trials supported the hypothesis that the CEMT could improve upon traditional security risk assessment

processes by providing decision-makers with the necessary understanding to make informed risk decisions.

Building upon these promising preliminary results, a full implementation of the CEMT was trialed on a complex naval capability. The findings of this implementation trial supported the hypothesis that the CEMT was an effective and efficient method for assessing the cyberworthiness of that capability. The research trial participants were in general agreement that the CEMT represents an improvement over existing approaches and facilitated informed and effective cyber security risk decisions.

The results for this research field trial provide important evidence to support a conclusion to our research questions. The results related to the effectiveness of the approach are not unanimous, but they are generally positive, giving us confidence that the CEMT can be an effective approach to assessing cyber-physical systems. With respect to the research question on the efficiency of the approach, there were two survey question responses that suggest that the CEMT models are not easy to produce and that the approach is still reliant on subjective data points. Notwithstanding, the results also indicated that there was reasonable success in being able to effectively inform decision-makers.

Broadly, the research field trial supports the following conclusions to each research question:

1. The CEMT can be a reasonably effective approach for assessing the cyberworthiness of real-world cyber physical systems.
2. The CEMT can be a somewhat efficient approach for assessing the cyberworthiness of real-world cyber physical systems, although much of this efficiency is gained to the ongoing evergreen usage of the tool and the ability to reuse large sections of the model on new systems.
3. The CEMT approach was able to generate outputs that meaningfully informed the decision-maker of the cyber risk within the system.

The research field trial also reconfirmed that the CEMT can meet all five key attributes of an effective model-based security assessment.

While this research field trial provides some confidence that the CEMT is an appropriate tool for conducting cyberworthiness assessments within the context of an Australian naval platform, there is opportunity for future research to investigate the application of the CEMT within the context of other complex cyber-physical systems. A single research trial cannot confirm the broad applicability of the CEMT, but the results of this research trial provide a strong impetus to undertake further research trials to validate the generalized applicability of the approach.

There are also opportunities for further development of the CEMT through the creation and refinement of reusable threat sets. These threat sets could be created through the application of the CEMT to real world systems, or even by applying the CEMT to a theoretical example system. The development of these threat sets could be combined into a centralised threat library that can be used as a starting point for ongoing assessments using the toolkit. Published threat libraries with stronger, public peer-review (aka crowd-surf) could help to address the latent subjectivity concerns identified in our implementation trial.

Finally, there are extensions to the CEMT itself that could be added to the toolkit as future work. The CEMT currently enumerates the actions that an attacker needs to take to compromise a system and articulates the opportunities to detect those actions. The natural extension would be to design and implement the ability to model the response and recovery to those attacker actions once they are detected. This addition to the CEMT would help to provide better coverage of all pillars of the NIST Cybersecurity Framework [31].

Our research has shown that the CEMT provides a sound basis for the assessment of cyberworthiness of real-world cyber-physical systems, and the identified future work could help to enhance and expand the applicability of the approach. This research is the first known research assessing the utility of model-based security assessment approaches (see Table 1) in a formal work-based assessment. Other researchers are encouraged to continue their research into field trials when expanding the real-world application of model-based security assessment approaches beyond the scope of the research trials described in this paper.

8 NOTES AND FUNDING

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

9 DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the authors used Grammarly in order to fix grammar mistakes. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

10 CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Stuart Fowler: Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Conceptualisation. Keith Joiner: Writing – review and editing, Resources, Methodology, Formal analysis, Conceptualisation, Supervision. Siqi Ma: Writing – review, Resources, Supervision.

11 REFERENCES

1. Australian Senate, Budget hearings on Foreign Affairs Defence and Trade, testimony by Vice Admiral Griggs, Major General Thompson and Minister of Defence, 29 May 2018, 20:33:25 – 20:34:52. Available: https://www.aph.gov.au/News_and_Events/Watch_Read_Listen/ParlView/video/691458 Accessed on 31 Oct 2024
2. Australian National Audit Office (ANAO), Defence's Management of ICT Systems Security Authorisations, Auditor General Report No. 2 2024-25. Available https://www.anao.gov.au/work/performance-audit/defences-management-of-ict-systems-security-authorisations Accessed on 31 Oct 2024
3. National Institute of Standards and Technology (NIST) Computer Security Resource Center (CSRC), Cyber Resiliency Definition. Available online: https://csrc.nist.gov/glossary/term/cyber_resiliency Accessed on 31 Oct 2024
4. National Institute of Standards and Technology (NIST) Computer Security Resource Center (CSRC), Cyber Survivability Definition. Available online: https://csrc.nist.gov/glossary/term/cyber_survivability Accessed on 31 Oct 2024
5. Purton, L., Kourousis, K. Military Airworthiness Management Frameworks: A Critical Review. Procedia Engineering, 2014, Volume 80, pp. 545–564.
6. Mo, J.P.T., Downey, K. System Design for Transitional Aircraft Support. International Journal of Engineering Business Management, 2014, Volume 6, pp. 45-56.
7. Hodge, R.J.,; Craig, S., Bradley, J.M., Keating, C.B. Systems Engineering and Complex Systems Governance – Lessons for Better Integration. INCOSE International Symposium, 2019, pp. 421–433.
8. Simmonds, S; Cook, S.C. Use of the Goal Structuring Notation to Argue Technical Integrity. INCOSE International Symposium;, 2017, pp. 826–841.
9. Australian Government. Defence Seaworthiness Management System Manual. 2018. Available online: https://www.defence.gov.au/sites/default/files/2021-01/SeaworthinessMgmtSystemManual.pdf (accessed on 15th of June 2024)
10. U.S. Department of Navy; Cyber Strategy, November 2023. Available online: https://dvidshub.net/r/irstzr (accessed on 25 April 2024).
11. Fowler S, Sitnikova E. Toward a framework for assessing the cyber-worthiness of complex mission critical systems. Military Communications and Information Systems Conference; 12-14 November; Canberra, Australia 2019.
12. Fowler, S., Joiner, K., & Ma, S. Cyber Evaluation and Management Toolkit (CEMT): Face Validity of Model-Based Cybersecurity Decision Making. Systems (Basel), 2024, Volume 12(7), pp. 238.
13. Cyber Evaluation and Management Toolkit (CEMT). Available online: https://github.com/stuartfowler/CEMT Accessed on 31 Oct 2024
14. Carter, B., Adams, S., Bakirtzis, G., Sherburne, T., Beling, P., Horowitz, B. A preliminary design-phase security methodology for cyber–physical systems. Systems (Basel), 2019, Volume 7(2), pp.21.
15. Larsen, M.H., Muller, G., Kokkula, S. A Conceptual Model-Based Systems Engineering Method for Creating Secure Cyber-Physical Systems. INCOSE International Symposium. 2022, pp. 202–213.
16. Larsen, M. H., Kokkula, S., & Muller, G. A Proposal for Model-Based Systems Engineering Method for Creating Secure Cyber-Physical Systems. INCOSE International Symposium. 2024, pp. 37–52.
17. Japs, S; Anacker, H., Dumitrescu, R. SAVE: Security & safety by model-based systems engineering on the example of automotive industry. 31st CIRP Design Conference, 2021.
18. Navas, J., Voirin, J., Paul, S., Bonnet, S. Towards a model-based approach to systems and cybersecurity: Co-engineering in a product line context. Insight (International Council on Systems Engineering). 2020, Volume 23, pp. 39–43.
19. Geismann, J., Gerking, C., Bodden, E. Towards ensuring security by design in cyber-physical systems engineering processes. International Conference on the Software and Systems Process. 2018, Gothenburg, Sweden.

20. Mažeika, D., Butleris, R. MBSEsec: Model-based systems engineering method for creating secure systems. Applied Sciences. 2020, Volume 10(7), pp. 2574.
21. Object Management Group. UAF: Unified architecture framework. 2022. Available online: https://www.omg.org/spec/UAF. Accessed on 31 Oct 2024
22. Jurjens, J. Secure Systems Development with UML. Berlin, Heidelberg: Springer; 2005.
23. Apvrille, L., Roudier, Y. Towards the model-driven engineering of secure yet safe embedded systems. International Workshop on Graphical Models for Security. 2014.
24. Verma, D. Systems Engineering for the Digital Age: Practitioner Perspectives (1st ed.). 2023, Wiley.
25. Silawi, E., Shaked, A., & Reich, Y. Translating the STPA-Sec Security Method into a Model-based Engineering Approach. INCOSE International Symposium, 2024, Volume 34(1), pp. 1948–1963.
26. Li, Y., Huang, C., Liu, Q., Zheng, X., & Sun, K. Integrating security in hazard analysis using STPA-Sec and GSPN: A case study of automatic emergency braking system. Computers & Security, 2024, Volume 142, pp. 103890.
27. MITRE ATT&CK Framework. Available online: https://attack.mitre.org/ (accessed on 25 April 2024).
28. Australian Government – Australian Signals Directorate, Information Security Manual (ISM). Available online: https://www.cyber.gov.au/resources-business-and-government/essential-cyber-security/ism Accessed on 31 Oct 2024.
29. National Institute of Standards and Technology (NIST) Computer Security Resource Center (CSRC), NIST Special Publication 800-53 Rev. 5: Security and Privacy Controls for Information Systems and Organizations. Available online: https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final Accessed on 31 Oct 2024
30. USDoD Cyber Table Top Guide, https://www.cto.mil/wp-content/uploads/2023/06/DoD-Cyber-Table-Top-Guide-v2-2021.pdf
31. National Institute of Standards and Technology (NIST), CSF 2.0: Cybersecurity Framework. Available online: https://www.nist.gov/cyberframework Accessed on 31 Oct 2024.

CHAPTER 7 CONCLUSION

7.1 OVERVIEW

This research project builds upon the work of the many other researchers working within the field of model-based cyber risk assessments, but the CEMT developed through this research project distinguishes itself from that existing work in the field by incorporating advanced graph-based techniques, the threat scenario vignettes from mission-based assessments and addressing the foundational cyberworthiness principles.

The CEMT also provides an efficient approach that incorporates all five key attributes that contribute to an effective model-based cyber risk assessment – an extended model-based taxonomy, a threat focus, detailed adversary modelling, visualisation & simulation of threat and explicit traceability of threats – which no existing model-based cyber risk assessment approach achieves. The effectiveness and efficiency of the CEMT was also demonstrated through practical implementation trials on complex, real-world systems, which showed the ability of the CEMT to scale beyond test beds and idealised theoretical systems.

There remain some limitations in the CEMT that must be addressed through future work, particularly with respect to incorporating response and recovery actions into the detailed threat model and providing validation of the expected reusability and maintainability of the tool through longitudinal studies. Additionally, opportunities exist to enhance the CEMT through the development of community threat sets and incorporating intelligent automation into the process.

In this chapter, the contributions made through this research project will be identified, each of the research questions will be individually discussed and answered, the overall significance of the research will be explored, and the open challenges and potential future work related to the CEMT will be described in more detail. This will be followed by concluding remarks that summarise the holistic research effort.

7.2 CONTRIBUTIONS

The contributions provided by this research project fall into two broad categories, the creation and development of a novel toolkit for performing cyber risk assessments and the collection of data from real-world implementation trials into the effectiveness and efficiency of that toolkit when applied to a complex cyber-physical system. The resultant contribution is the advancement of professional cybersecurity practice through the provision of an implementable toolkit that can be applied by other practitioners within the field while also demonstrating the efficacy of that toolkit during real-world applications.

The first category of contribution is the development of the Cyber Evaluation and Management Toolkit (CEMT) itself. This toolkit implements a novel methodology for completing threat-focused cyber risk assessments of cyber-physical systems using model-based systems engineering design techniques. This allows for detailed threat modelling of holistic systems-of-systems in a manner that supports the efficient ongoing sustainment and maintenance of a living cyber risk assessment, as well as the rapid reuse of those threat models in future assessments of similar systems in similar threat environments.

The toolkit provides strong automation, both during threat model development and peer review, and transparent traceability that leverages the inherited capabilities of model-based systems engineering tools. The CEMT has been developed and published under an open-source license that allows the cybersecurity community to freely use the toolkit for their own commercial and

non-commercial purposes [118]. This, on its own, is a significant contribution to professional cybersecurity practice.

However, this research project also provides cybersecurity professionals insight, based on practical trials, into how well the threat-focused methodology implemented within the CEMT can evaluate the cyberworthiness of cyber-physical systems in terms of both efficiency and effectiveness. The research also explored whether the CEMT helped to achieve informed decision-making and compared its threat-focused approach to the prevailing approaches within industry.

7.2.1 Implementation Trial Results Summary

Chapter 6 provided a detailed description of the practical trial of the CEMT and included a set of results from that trial, which was conducted on a large naval capability, including the IT, OT and integrated cyber-physical systems within that capability. These results were the responses to the survey questions outlined in Chapter 3. The overall response to the CEMT was broadly very positive, with 62.5% of respondents saying that the CEMT was either Excellent or Very Good as a cyberworthiness assessment methodology and when respondents were asked to rate how likely they were to recommend the CEMT for future projects on a scale of 0-10 the average of responses was 8.6.

To aid with the discussion of contributions towards answering the overall research questions, the results from the implementation trial are summarised in

the following tables which aggregates the 5-point Likert scale used in the survey questionnaire into a 3-point scale that shows broad agreement, disagreement and neutrality by combining the *Agree* and *Strongly Agree* responses into a single *Agree* metric and combining the *Disagree* and *Strongly Disagree* metric into a single D*isagree* metric. These tables have been categorised based on how positive the responses were to each question, with Table 7-1 showing the questions which had an overwhelming positive response where more than 80% of respondents agreed or strongly agreed with the survey question. The items in Table 7-1 therefore represent questions where the cohort of respondents were very aligned in their agreement with the statements.

Table 7-1: *Survey questions with overwhelming positive responses*

Survey Question	Agree	Disagree	Neutral
Q9: Facilitates Meaningful Stakeholder Engagement	97%	0%	3%
Q13: Identifies Why Controls Are Important	97%	0%	3%
Q17: Identifies Additional Controls	97%	0%	3%
Q22: Provides Ongoing Value	97%	0%	3%
Q34: Easy to Update as Design Matures	97%	0%	3%
Q37: Adds Value to a Project	97%	0%	3%
Q39: Cyberworthiness is Critical	97%	0%	3%
Q2: Encourages Collaboration	94%	3%	3%
Q6: Identifies Factors Contributing to Cyberworthiness	91%	0%	9%
Q18: Identifies Relative Impact of Controls	91%	3%	6%
Q19: Produces Transparent Assessments	91%	0%	9%
Q29: Facilitates Effective Decisions	91%	0%	9%
Q32: Adaptable to System Changes	91%	3%	6%
Q36: Supports Evergreen Development	91%	3%	6%
Q38: Improvement Over Existing Approaches	91%	0%	9%
Q11: Facilitates Justification of Recommendations	88%	3%	9%
Q12: Identifies Important Controls	88%	0%	12%
Q20: Tailored to System Context	88%	6%	6%
Q23: Highlights Areas of Risk	88%	3%	9%
Q26: Facilitates Informed Decisions	88%	0%	12%
Q28: Provides New Insights	88%	3%	9%
Q33: Adaptable to Threat Changes	88%	3%	9%
Q5: Sufficiently Detailed	84%	0%	16%
Q21: Promotes Collaboration	84%	3%	13%

Survey Question	Agree	Disagree	Neutral
Q35: Agile and Responsive	84%	3%	13%
Q40: Improves Understanding of System Cyberworthiness	84%	0%	16%
Q1: Scoped to Context	81%	6%	13%
Q14: Understandable Risk Presentation	81%	0%	19%
Q30: Supports Accountability	81%	6%	13%

Table 7-2 displays those questions that had broadly positive responses, but there was some uncertainty in the responses. These questions had less than 80% of responses as agree or strongly agree but still had less than 10% disagree or strongly disagree. These items in Table 7-2 are those questions where the group of respondents were broadly in agreement with the statement, but a significant proportion of respondents with neutral in their response. This likely indicates that there was some uncertainty for a significant number of respondents, and they would possibly require further exposure to the CEMT for an opinion to form either way.

Table 7-2: *Survey questions with positive responses with some uncertainty*

Survey Question	Agree	Disagree	Neutral
Q7: Effective Use of Time	78%	3%	19%
Q31: Presents Information Effectively	78%	0%	22%
Q24: Identifies Why Decisions Are Made	75%	3%	22%
Q25: Identifies How Decisions Are Made	75%	0%	25%
Q27: Complete & Comprehensive	72%	6%	22%
Q8: Intuitive	69%	3%	28%
Q16: Easy to Understand	69%	0%	31%
Q15: Produces Accurate Assessments	66%	3%	31%
Q41: Appropriate Cyberworthiness Technique	66%	3%	31%
Q4: Good Balance Between Quality and Time Taken	59%	3%	38%

Table 7-3 highlights those questions that had mixed responses from the respondents, where less than 50% of respondents agreed or strongly agreed with the statement and more than 15% of respondents disagreed or strongly disagreed with the statement. This likely indicates that the respondents were split

in their opinions on these questions and ultimately the results of the implementation trial are inconclusive for these statements and therefore they cannot be confirmed as attributes or outcomes of the CEMT methodology.

Table 7-3: *Survey questions with mixed responses*

Survey Question	Agree	Disagree	Neutral
Q10: Not Overly Subjective	47%	22%	31%
Q3: Simple to Produce	41%	18%	41%

The summarised results outlined in these tables will be referred to within the following sections to support the discussion and conclusions with respect to each of the research questions defined in Section 3.2.

7.2.2 Primary Research Question

The primary research question for this research project is:

> *Are integrated threat models, developed using Model-based Systems Engineering tools, an effective and efficient basis for the assessment and evaluation of cyberworthiness in cyber-physical systems?*

This research question focuses on two primary measures, the effectiveness of the CEMT at evaluating cyberworthiness and the efficiency of the CEMT when conducting that evaluation. The results of the primary implementation trial showed that the integrated threat models developed using the CEMT were able to effectively assess the cyberworthiness of a cyber-physical system. The overall effectiveness of the methodology at performing cybersecurity risk assessments in general was supported in the data in three categories:

1. Identified Risks and Mitigations; supported by the overwhelming positive responses to the following questions:
    - Q12: Identifies important controls
    - Q17: Identifies additional controls
    - Q18: Identifies relative impact of controls
    - Q23: Highlights areas of risk
2. Level of Detail; supported by the overwhelming positive responses to the following questions:
    - Q5: Sufficiently detailed
    - Q11: Facilitates justification of recommendations
    - Q13: Identifies why controls are important
3. Useful Outputs; supported by the overwhelming positive responses to the following questions:
    - Q28: Provides new insights
    - Q37: Adds value to a project
    - Q22: Provides ongoing value
    - Q6: Identifies factors contributing to cyberworthiness

The overwhelmingly positive responses to these questions provide strong support for a conclusion that the CEMT can effectively evaluate the cyberworthiness of a cyber-physical system. The methodology clearly identifies

the areas of risks and the key mitigations that are in-place, or could be put in-place, to control those risks. The assessment provides a sufficient level of detail to support risk analysis and produces outputs that are insightful and identify the key factors that contribute to cyberworthiness.

However, there was an outlier in the data related to the effectiveness of the CEMT, which was *Q10: Not Overly Subjective*. This question received mixed responses, and while the responses were still more positive than negative, this result highlights that the CEMT process is still reliant on the subjective opinions of subject matter experts. This is an insightful result, as the CEMT does not change the way in which cyber risk assessments are performed, as much as it changes the way that those cyber risk assessments are documented and reviewed.

As discussed in Section 1.2.3, cybersecurity is an inherently adversarial domain where information is being deliberately obfuscated by that adversary, and because of this there is uncertainty in the analysis that requires a risk-based approach that doesn't have access to perfect input data. This unknown or imperfect input data forces subjectivity into the analysis, and the CEMT does not overcome that constraint.

The results of the trial also showed that efficiency was demonstrated in three different ways, specifically:

1. Efficient initial development; supported by the positive responses with some uncertainty to the following questions:
    - Q7: Effective use of time
    - Q4: Good balance between quality and time taken
2. Minimisation of Nugatory Effort; supported by the overwhelming positive responses to the following questions:
    - Q1: Scoped to context
    - Q20: Tailored to system context
3. Efficiency through Maintainability; supported by the overwhelming positive responses to the following questions:
    - Q32: Adaptable to system changes
    - Q33: Adaptable to threat changes
    - Q34: Easy to update as design matures
    - Q35: Agile and responsive
    - Q36: Supports evergreen development

The results for the efficiency of the initial development of the threat model are not as clear as the results were for the effectiveness of the methodology. This is underscored by the responses for *Q3: Simple to Produce*, which had the least positive responses of any question asked in the survey. This result reflects the fact that the CEMT requires the development of a detailed threat model which

does involve activities that are neither quick nor simple. However, efficiency isn't just about doing things quickly, as it is a measure of productivity, not just speed. In this context, the positive responses for *Q7: Effective use of time* and *Q4: Good balance between quality and time taken* provide comfort that a measure of efficiency was still gained through the initial development using the CEMT.

The results for efficiency through the minimisation of nugatory efforts were more compelling. By effectively scoping the assessment to just the context of the system being evaluated and only deriving controls that are directly contributing to the mitigation of relevant threats to that system context, the assessment efforts are properly focused onto areas that add value. This ability to avoid the assessment and consideration of mitigations that are not relevant to the system context adds to the efficiency of the cyberworthiness assessment.

In combination with the efficiency gains through maintainability, which are discussed in more detail in the exploration of research sub-question 2 in Section 7.2.4, these positive responses to the efficiency questions within the survey provide strong support for a conclusion that the CEMT can efficiently evaluate the cyberworthiness of a cyber-physical system.

In addition to the effectiveness and efficiency of the CEMT process, the trial results also allow us to explore some of the assurance concepts related to cyberworthiness. As identified in Section 3.1.3 the key principles that drive cyberworthiness relate to the way in which the assessment is performed and how

that influences the understanding of stakeholders involved in the process, specifically it requires that decisions are:

1. Mindful – the operating context of the system is thoroughly understood during the assessment process
2. Collaborative – all relevant stakeholders engage in joint problem solving during the assessment process
3. Transparent – everyone involved in the assessment process understands what was decided and why
4. Accountable – decision makers are sufficiently informed for them to be accountable for the decisions they make in the assessment process

The results of the implementation trial confirmed that the CEMT encouraged mindfulness from the assessment stakeholders throughout the evaluation process, which is supported by the overwhelmingly positive responses to the *Q1: Scoped to context* and *Q20: Tailored to system context* questions. Respondents clearly agreed that the CEMT process provided a strong and recurrent focus on the context of the system, facilitating mindful analysis.

Collaboration amongst the assessment stakeholders was also supported by the CEMT, as shown by the positive responses to *Q21: Promotes collaboration*, *Q2: Encourages collaboration* and *Q9: Facilitates Meaningful Stakeholder Engagement*. This effective collaboration ensured that the relevant stakeholders were able to support the cyberworthiness analysis and evaluation through joint problem solving.

The transparency of the methodology was also demonstrated through the practical trial, as evidenced by the positive responses to Q19: Produces transparent assessments, Q11: Facilitates justification of recommendations, Q24: Identifies how decisions are made and Q25: Identifies why decisions are made. This transparency ensures that all stakeholders are aware of what is being described by the analysis, which allows for sensible and efficient peer review.

Finally, the accountability of decision makers was also found the be supported by the CEMT, as seen in the overwhelmingly positive responses to Q30: Supports accountability. The detail behind the ways in which informed decision making was evaluated is discussed further in the exploration of research sub-question 1 in Section 7.2.3.

Given this data, we can conclude that yes, integrated threat models, developed using Model-based Systems Engineering tools, are an effective and efficient basis for the assessment and evaluation of cyberworthiness in cyber-physical systems. Answering this primary research question contributes to professional cybersecurity practice by demonstrating through an initial trial that the CEMT can be both effective and efficient at evaluating the cyberworthiness of real-world cyber-physical systems. While this is only a single trial and cannot be easily generalised across all cyber-physical systems, the results provide confidence that the toolkit is implementable under realistic conditions and can efficiently produce effective cyberworthiness evaluations. This practical evidence is critical to cybersecurity practitioners evaluating cyber-physical systems, as the potential

consequences of applying an untested methodology can dissuade innovative academic methods from being applied in practice.

7.2.3 Research Sub-Question 1

The first research sub-question for this research project is:

> *Do the developed threat models provide decision makers with the necessary understanding to make informed security risk decisions?*

This research question focuses on the use of threat models to convey sufficient understanding of the threat environment, system context and threat vectors to decision makers for them to make informed decisions about the acceptability of any residual risks. Understanding, in this context, can be broken down into two categories, understanding the analysis performed within the CEMT and understanding the output of the CEMT.

The understandability of the analysis performed within the CEMT is best highlighted by the positive responses to the following survey questions during the implementation trial:

- Q8: Intuitive
- Q16: Easy to understand
- Q31: Presents information effectively

While there was some uncertain amongst the respondents with respect to these questions, the overall response was still broadly positive, which provides

reasonable confidence that the CEMT supports an understanding of the cyberworthiness analysis performed within the tool without requiring high level of cybersecurity expertise.

The understandability of the outputs of the CEMT, which are the primary interaction points with decision makers, is demonstrated through the overwhelmingly positive responses to the following questions:

- Q6: Identifies factors contributing to cyberworthiness
- Q14: Understandable risk presentation
- Q29: Facilitates effective decisions
- Q40: Improves understanding of system cyberworthiness

The responses to these questions make it clear that the CEMT can present risk to decision makers in an understandable manner and improves the understanding of the cyberworthiness of the system by clearly identifying the key factors that contribute to the system's cyberworthiness to those decision makers. Ultimately, this allows for informed and effective risk decisions to be made with respect to the cyberworthiness of a system.

Given this data, we can conclude that yes, the threat models, developed using Model-based Systems Engineering tools, do provide decision makers with the necessary understanding to make informed security risk decisions. Answering this research sub-question contributes to professional cybersecurity practice by demonstrating through a practical trial with real-world executive decision makers

that the CEMT can effectively communicate detailed cybersecurity analysis to those decision makers in an understandable format, facilitating informed decision makers that can reasonably be expected to accept accountability for those decisions.

7.2.4 Research Sub-Question 2

The second research sub-question for this research project is:

> *Does the process provide sufficient reusability and maintainability that the methodology is more efficient than prevailing control-focused approaches?*

This research question focuses on ability for the efficiency benefits of the threat modelling process over the life of a system due to improved maintainability of the assessment and across a family of systems through reuse of previous assessments.

The ability to efficiently maintain the assessment is captured in the overwhelmingly positive responses to the following survey questions asked during the implementation trial:

- Q32: Adaptable to system changes
- Q33: Adaptable to threat changes
- Q34: Easy to update as design matures
- Q35: Agile and responsive

- Q36: Supports evergreen development

This data indicates that the respondents felt that the CEMT was able to easily incorporate updates to the analysis caused by changes to the threat environment or system design. The CEMT methodology provided a level of agility in the cyberworthiness assessment that allowed it to remain a living assessment to support an evergreen development cycle for the system being evaluated, which provides a strong indication that the guiding design principle for maximising maintainability and reusability used during development of the CEMT has been followed.

However, it must be noted that the survey for the implementation trial captured data at a single point in time towards the end of the trial. Consequently, the respondents would be familiar with the ability for the CEMT to incorporate changes that occurred during the trial period as well as the reusability features within the tool and could apply that knowledge to predict how the CEMT could be updated and reused over time, but they did not experience a long-term implementation of the CEMT or an implementation of the CEMT over multiple systems.

This means that valid, comprehensive conclusions related to the maintainability and reusability of the CEMT process cannot be drawn based on the data gathered from this research project alone. The findings of the implementation trial, as well as the discussion of CEMT threat sets in Section 4.2.4, support a strong hypothesis that the CEMT could provide significant efficiency gains due to

maintainability and reusability, but formally testing and confirming that hypothesis would require longitudinal studies using the CEMT, which is left as an open challenge and discussed as potential future work in Section 7.4.3.

7.2.5 Research Sub-Question 3

The final research sub-question for this research project is:

> *Do cyber security risk practitioners prefer the integrated threat model approach to traditional security risk assessment processes for assessing cyberworthiness?*

This research question focuses on a direct comparison between the integrated threat modelling approach of the CEMT and the contemporary security risk assessment processes described in Section 1.2.1. This is primarily a measure of user satisfaction with the threat modelling process, compared to the existing experiences with the contemporary approaches.

This preference is directly measured using the survey question Q38: Improvement over existing approaches, which explicitly asked respondents whether they agreed that the CEMT is an improvement over existing cyberworthiness assessment processes. 91% of respondents either agreed or strongly agreed with that statement, and no respondents either disagreed or strongly disagreed with that statement.

While the practical trial was just a single implementation with a single set of cyber security risk practitioners participating in the trial and it is difficult to generalise these responses universally, the overwhelming positive response does provide us with the confidence to conclude that practitioners do prefer the integrated threat modelling approach provided by the CEMT over the contemporary approaches for assessing cyberworthiness. This is a strong indication that the CEMT does provide a significant contribution to the professional cybersecurity practice, as members of that professional practice have a reported preference for using the CEMT over contemporary alternatives currently used within industry.

7.3 SIGNIFICANCE

Answering the research questions is a critical outcome for the research project, but it is also important to address the original research goal and motivations behind that research goal. In this section, the motivations for the research and the resultant research goal will be revisited and discussed in the context of the completed research project.

The research goal was defined in Section 1.4 and is copied below:

> *The explicit goal of the research is therefore to investigate threat-focused approaches that leverage model-based analysis techniques to assess the cyberworthiness of cyber-physical systems, with a particular focus on the ability of those approaches to facilitate informed decision making with respect to cybersecurity risk acceptance.*

The research project was able to achieve this goal and the results of the practical implementation trial using the CEMT show a successful investigation of threat-focused approaches that leverage model-based systems engineering techniques to assess the cyberworthiness of a cyber-physical system. Through a focus on the cyberworthiness principles and the gathering of data from decision makers involved with the trial, we were able to place a focus on the ability of the threat-focused process to facilitate informed decision making.

The broader motivation for the research, as detailed in Section 1.2, was about the inability of the contemporary approaches to cybersecurity risk assessment to achieve this research goal. As summarised in Section 7.2.5, the results of the implementation trial showed that there was very strong agreement amongst trial participants that the CEMT approach was an improvement over contemporary approaches when it comes to assessing the cyberworthiness of cyber-physical systems. These results provide confidence that the research project has successful helped to address the original motivation for conducting the research.

While the importance of achieving this motivation was already outlined in Section 1.2.3, it is worth acknowledging that the participants in the implementation trial agreed that addressing the cyberworthiness of cyber-physical systems was critical. This is demonstrated by 97% of respondents either agreeing or strongly agreeing with *Q39: Cyberworthiness is critical*, which asked participants whether they agreed that a comprehensive cyberworthiness assessment is a critical component of the system capability lifecycle. Clearly, the methodology developed

and tested during this research project has provided an important and significant contribution to professional cybersecurity practice.

The developed methodology is also adaptable and potentially applicable outside purely cyber-physical systems. The CEMT approach is not constrained to only cyber-physical systems and could also be applied to more traditional IT and OT systems if additional rigour and assurance is required when assessing those systems.

However, it is also important to note that the CEMT should not be considered a panacea for the assessment of the cyberworthiness of cyber-physical systems. The implementation trials did identify that there is still significant subjectivity in the assessments and that the threat models that form the basis of a CEMT assessment are not simple to produce. There may be systems that require even higher assurance and objectivity than is provided by the CEMT, and formal methods would likely be more appropriate for those systems. Equally, there may be less critical cyber-physical systems that don't need the rigour of the CEMT approach and therefore could be addressed through a simpler methodology. The CEMT does address that middle ground though, where decision makers and risk acceptance authorities need a level of assurance that exceeds that provided by contemporary approaches, but do not need the extreme rigour and objectivity provided by formal methods.

Of course, it is also important to acknowledge that the CEMT is not perfect at addressing that middle ground either. There remain several open challenges that

can be addressed through further research and development of the CEMT, and these will be discussed in the remainder of this chapter.

7.4 OPEN CHALLENGES AND FUTURE WORK

While significant progress has been made through this research, there were several limitations of the work that performed which provide numerous avenues for future work. The primary research limitations are:

1. The current iteration of the CEMT focuses on prevention and detection of threat actors, with no explicit methodology for addressing response to and recovery from a successful attack.
2. While the CEMT itself is open source and freely available, the models and threat sets built through the implementation trials have not been made public.
3. The implementation trials covered the initial creation of the threat model and a single point-in-time assessment of a cyber-physical system and therefore cannot provide strong validation of the efficacy of the CEMT's expected benefits for reusability and sustainability.
4. The selection of critical risk vectors to take forward to the risk assessment phase is manual and subjective, with no clearly defined criteria for which threat vectors should be selected for detailed quantitative analysis.

These limitations are left as open challenges for future research efforts to explore and address. In the following sections, each of these challenges will be explained in further detail to guide potential future researchers.

7.4.1 Recovery and Response

The hypothesis outlined in Section 3.1.3 provided a scope definition for Threat-based Cybersecurity Engineering by describing the inputs and outputs of the approach. The primary outputs were identified as:

- Preventative Controls – a baseline of preventative cybersecurity controls for inclusion in the cyber design of a system
- Detecting Controls – a baseline of detection and response controls for implementation by the Cyber Operations team
- Recovery Controls – a baseline of recovery and resilience controls for implementation by the System Operations team
- Residual Risk – the overall risk presented by the threats to the capability given the mitigation mechanisms that are in place

The relevant outputs here from the perspective of this open challenge is that 'response controls' phrase and the 'recovery and resilience controls' phrase under the *Detecting Controls* and *Recovery Controls* items, respectively. Clearly, there was an intent for a complete and comprehensive Threat-based Cybersecurity Engineering process to include an analysis and assessment of response and recovery activities related to the system being evaluated.

However, as described in Chapter 4, the CEMT focuses primarily on preventative controls, which are linked to `ThreatActions`, and detecting controls, which are linked to `DetectionActions`. There is no explicit modelling of response and recovery actions and therefore no explicit traceability and derivation of any associated response controls or recovery controls with the current CEMT. There are several ways in which this part of the Threat-based Cybersecurity Engineering scope could be incorporated into the CEMT, with two potential options discussed below.

Firstly, a detailed modelling approach could be taken, which parallels the approach taken for incorporating detection actions. In this approach, specific actions that represent the response and recovery events would be created for each `DetectionAction` in the model and they would articulate how cyber defenders could response to a detection and attempt to recover the system at that point in the threat vector. `SecurityControls` could then be linked to those response and recovery actions to identify controls that could be implemented to make that response and/or recovery possible.

Secondly, a less detailed modelling approach could be taken, where response and recovery controls are not explicitly linked to dedicated actions in the mal-activity diagrams during the *Threat Modelling* phase but are instead linked to overall security risks during the *Risk Assessment* phase. This could be achieved by linking response and recovery controls to the `ThreatImpact` object, which represents the final step in a threat vector, or they could be linked to the

`SecurityRisk` objects, which represents an entire end-to-end threat vector that is being quantitatively simulated.

There is expected to be an efficiency vs effectiveness trade-off between these two options, and future work would need to investigate these trade-offs, as well as comparing these potential options to other available options for addressing response and recovery controls. It may be that the CEMT is not the best way to derive these controls, and a separate but complementary process may be more effective and efficient. In any case, evaluating these pros and cons through implementation trials is left for future work to explore.

7.4.2 Community Threat Sets

One of the primary contributions of this research was to make the CEMT open-source and freely available for others to apply to their own cyber-physical systems or for future researchers to build upon. However, not everything created through the implementation trials in this research project has been made publicly available. Specifically, the potentially reusable threat set, which is described in Section 4.2.4, that was created through this research project describes cyber risks to a real-world system and consequently contains sensitive data that cannot be shared publicly.

What is available as part of the open-source CEMT is the profile which implements the methodology to create threat sets as well as a generic sample model that provides simple examples of how the CEMT can be applied. Of

course, having full threat sets available publicly would help others apply the CEMT to their system more efficiently, as they would have a starting point which could be directly applied to their cyber-physical system or modified to cover their particular use case.

While the tools to create a library of community threat sets for common cyber-physical system patterns is available through the open-source CEMT, the creation of these threat sets has not been completed as part of this research project and is left as an open challenge for future work. Researchers and community members looking to address this challenge could develop their own threat sets using the CEMT and then publish them for other community members to utilise.

7.4.3 Reusability and Maintainability

Another challenge that is left open following this research project is the effect that the expected reusability and maintainability of the CEMT assessment will have on the overall efficiency of cybersecurity assessments across an organisation. While there are hypothesised benefits to the overall efficiency at an organisational level to using model-based cybersecurity approaches, this research project only looked at a single assessment for a real-world system, and as such is not able to provide definitive conclusions around the ability for the CEMT to support this outcome.

To provide validity to the claim that a threat-focused, model-based approach to cyberworthiness assessments, such as the CEMT, would deliver efficiency benefits due to increased reusability across projects and maintainability of the assessment as there are changes in the threat environment or system design, additional research would need to be conducted. This would likely take the form of longitudinal studies that observed the rollout of a CEMT approach across an organisation, or the upkeep of a CEMT-based assessment over time.

To address this challenge, it is unlikely that additions or changes would need to be made to the CEMT. Future work to address this challenge would likely focus on experiment design and the development of a valid methodology for conducting the longitudinal studies required to capture the necessary data.

7.4.4 Intelligent Automated Analysis

The final open challenge relates to the ability to conduct a level of intelligent, automated analysis to assist modellers to make some of the subjective decisions that need to be made during the CEMT process. These decisions include:

- Review of system design to identify relevant threat paths and determine the appropriate threat sets.
- Selecting the critical threat vectors from the attack tree that will progress quantitative analysis during the *Risk Assessment* phase.

- Estimation of control effectiveness values based on control implementation states and an understanding of how the controls have been implemented

The identification of relevant threat paths through an automated review of the system design would streamline the process of threat modelling, increasing the efficiency of the overall CEMT process. One key component of this challenge is that the CEMT approach does not mandate a system design to be provided in a particular format. While this provides some flexibility in how the CEMT is applied, this flexibility can constrain the ability for an automated tool to conduct this assessment. However, if the system design was also provided in a SysML model, the design could be interpreted by a machine and intelligent automated analysis could be performed on that design to identify and build a threat model.

The selection of critical threat vectors from the attack tree is another open challenge that might be more achievable if the input of a system design in an SysML format were enforced. However, with sufficient data it may be possible to make sensible recommendations to the modeller about which threat vectors in the overall threat model should be selected for further analysis as a specific risk. This would help to provide confidence that the correct subset of threat vectors was selected for more detailed analysis.

Like the selection of critical threat vectors, the determination of control effectiveness is a subjective step in the CEMT approach. While the intention is for peer review to help ensure consistency and accuracy in these subjective

decisions, a future improvement could be to perform a level of automated intelligent analysis of the context of a particular step in the threat vector and the implementation state of controls tied to that step to estimate the effectiveness of those controls.

While this automated assessment may not be perfectly accurate, a reasonable estimation based on historical data would still provide significant value to the modelling process, in terms of both consistency and overall efficiency. Implementable solutions to these open challenges would likely have a noticeable impact on the efficiency and effectiveness of the CEMT approach and would consequently be a fertile area for future work to build upon this research project.

7.5 CONCLUDING REMARKS

The Cyber Evaluation and Management Toolkit (CEMT) developed and tested through this research project provides the professional cybersecurity practice with a novel methodology for the assessment of cyberworthiness within cyber-physical systems. This toolkit was shown to be both effective and efficient at performing cyberworthiness evaluations through practical trials on a real-world system.

Through this practical trial, professional cybersecurity practitioners evaluated the CEMT and agreed that the toolkit represented an improvement over contemporary approaches to the assessment of cyberworthiness in cyber-physical systems. The toolkit also facilitated the effective presentation of risk to

decision-makers, improving their understanding of the cybersecurity risks associated with the system and the key factors driving those risks, which contributed to more informed cybersecurity risk decisions being made.

The research project demonstrated that the CEMT is an implementable methodology that can perform cybersecurity risk assessments at a system-of-systems level to provide actionable insights into the holistic cyberworthiness of a capability. Overall, this research project represents a significant contribution to professional cybersecurity practice through both creation and free distribution of the CEMT and the gathering of practical data on the CEMT's efficacy when applied to complex cyber-physical systems in the real world.

However, this version of the CEMT is just the starting point for the ongoing development for model-based, threat-focused approaches to the cyberworthiness assessment of cyber-physical systems. While the promise of threat-focused, model-based approach to assessing the cyberworthiness of cyber-physical systems to facilitate informed decision-making has been demonstrated through this initial research and development activity, there are still numerous open challenges that future research and development activities can address by building upon the foundation provided by the CEMT. Hopefully the results of this research project can inspire future researchers to take on those challenges and further enrich the literature through their own research.

REFERENCES

1. Kale, V. (2015). Networking and Internetworking. In *Guide to Cloud Computing for Business and Technology Managers* (pp. 66–99). Chapman and Hall/CRC. https://doi.org/10.1201/b17923-6
2. Campbell-Kelly, M., Aspray, W. F., Yost, J. R., Tinn, H., Con Díaz, G., & Ensmenger, N. (2023). New Ways of Computing. In *Computer* (4th ed., pp. 209–231). Routledge. https://doi.org/10.4324/9781003263272-13
3. Campbell-Kelly, M., Aspray, W. F., Yost, J. R., Tinn, H., Con Díaz, G., & Ensmenger, N. (2023). The Shaping of the Personal Computer. In *Computer* (4th ed., pp. 235–255). Routledge. https://doi.org/10.4324/9781003263272-15
4. Kozierok, C. M. (2005). *The TCP/IP guide : a comprehensive, illustrated Internet protocols reference* (1st ed.). No Starch Press.
5. Stouffer K, Pease M, Tang CY, Zimmerman T, Pillitteri V, Lightman S, Hahn A, Saravia S, Sherule A, Thompson M (2023) Title. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 800-82r3. https://doi.org/10.6028/NIST.SP.800-82r3
6. Manesis, S., & Nikolakopoulos, G. (2018). *Introduction to Industrial Automation* (First edition.). CRC Press. https://doi.org/10.1201/9781351069083
7. Gibert, D., Mateu, C., & Planes, J. (2020). The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. *Journal of Network and Computer Applications*, *153*, 102526-. https://doi.org/10.1016/j.jnca.2019.102526
8. Liska, A. (2015). *Building an intelligence-led security program* (T. Gallo, Ed.; First edition.). Syngress.
9. Newsweek (2004). More Doom? Available Online: https://www.newsweek.com/more-doom-131157 Accessed March 2025
10. Knapp, E. D. (2024). Industrial Cybersecurity History and Trends. In *Industrial Network Security*. Elsevier Science & Technology Books.
11. National Health Service (2023). NHS England business continuity management toolkit case study: WannaCry attack. Available Online: https://www.england.nhs.uk/long-read/case-study-wannacry-attack/ Accessed March 2025
12. WIRED (2018). The Untold Story of NotPetya, the Most Devastating Cyberattack in History. Available Online: https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/ Accessed March 2025
13. Australian Department of Home Affairs, PSPF Annual Release, Available Online: https://www.protectivesecurity.gov.au/pspf-annual-release Accessed March 2025
14. Australian Government. Information Security Manual. Available Online: https://www.cyber.gov.au/resources-business-and-government/essential-cybersecurity/ism Accessed March 2025
15. Australian Government. Essential Eight Maturity Model. Available Online: https://www.cyber.gov.au/resources-business-and-government/essential-cyber-security/essential-eight/essential-eight-maturity-model Accessed March 2025
16. Australian Government. Essential Eight Explained. Available Online: https://www.cyber.gov.au/sites/default/files/2023-

11/PROTECT%20-%20Essential%20Eight%20Explained%20%28November%202023%29.pdf Accessed March 2025
17. US Government. Federal Information Security Modernization Act of 2014. Available Online: https://www.congress.gov/113/plaws/publ283/PLAW-113publ283.pdf Accessed March 2025
18. National Institute of Standards and Technology. Special Publication 800-37 Rev 2: Risk Management Framework for Information Systems and Organizations. Available Online: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-37r2.pdf Accessed March 2025
19. National Institute of Standards and Technology. Special Publication 800-53 Rev 5: Security and Privacy Controls for Information Systems and Organizations. Available Online: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r5.pdf Accessed March 2025
20. Purton, L., & Kourousis, K. (2014). Military Airworthiness Management Frameworks: A Critical Review. *Procedia Engineering*, *80*, 545–564. https://doi.org/10.1016/j.proeng.2014.09.111
21. Mo, J. P. T., & Downey, K. (2014). System Design for Transitional Aircraft Support. *International Journal of Engineering Business Management*, *6*(Godište 2014), 6-. https://doi.org/10.5772/57451
22. Hodge, R. J., Craig, S., Bradley, J. M., & Keating, C. B. (2019). Systems Engineering and Complex Systems Governance – Lessons for Better Integration. *INCOSE International Symposium*, *29*(1), 421–433. https://doi.org/10.1002/j.2334-5837.2019.00612.x
23. Simmonds, S., & Cook, S. C. (2017). Use of the Goal Structuring Notation to Argue Technical Integrity. *INCOSE International Symposium*, *27*(1), 826–841. https://doi.org/10.1002/j.2334-5837.2017.00396.x
24. National Institute of Standards and Technology. OSCAL: the Open Security Controls Assessment Language. Available Online: https://pages.nist.gov/OSCAL/ Accessed March 2025
25. Henderson, K., & Salado, A. (2021). Value and benefits of model-based systems engineering (MBSE): Evidence from the literature. *Systems Engineering*, *24*(1), 51–66. https://doi.org/10.1002/sys.21566
26. Australian National Audit Office (2024). Defence's Management of ICT Systems Security Authorisations. Available Online: https://www.anao.gov.au/work/performance-audit/defences-management-of-ict-systems-security-authorisations Accessed March 2025
27. U.S. Department of Navy; Cyber Strategy, November 2023. Available Online: https://dvidshub.net/r/irstzr Accessed March 2025
28. Griffor, E. , Greer, C. , Wollman, D. and Burns, M. (2017), Framework for Cyber-Physical Systems: Volume 1, Overview, Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, [online], https://doi.org/10.6028/NIST.SP.1500-201 (Accessed March 3, 2025)
29. Nardelli, P. H. J. (2022). *Cyber-physical systems : theory, methodology, and applications*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781119785194

30. Pathan, A.-S. K. (Ed.). (2016). Securing cyber-physical systems. CRC Press, Taylor and Francis Group.
31. Nand, P. (Ed.). (2022). *Cyber-physical systems : foundations and techniques*. John Wiley & Sons. https://doi.org/10.1002/9781119836636
32. Rawat, D. B., Rodriques, J., & Stojmenovic, I. (Eds.). (2016). *Cyber physical systems : from theory to practice* (1st edition). Taylor & Francis. https://doi.org/10.1201/b19290
33. Fowler, S., Sweetman, C., Ravindran, S., Joiner, K., & Sitnikova, E. (2017). *Developing cyber-security policies that penetrate Australian defence acquisitions*.
34. Australian Senate. Budget Hearings on Foreign Affairs Defence and Trade, Testimony by Vice Admiral Griggs, Major General Thompson and Minister of Defence (29 May, 2033-2035 hours). 2018. Available Online: https://parlview.aph.gov.au/mediaPlayer.php?videoID=399539 timestamp 3:19:43. Access March 2025
35. Australian Government. ADF Cyberworthiness Governance Framework. Canberra, 2020.
36. Swanson, E. B., & Culnan, M. J. (1978). Document-Based Systems for Management Planning and Control: A Classification, Survey, and Assessment. *MIS Quarterly*, *2*(4), 31–46. https://doi.org/10.2307/248903
37. Weilkiens, T. (2022). *Model-based system architecture* (Second edition.). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119746683
38. Feiler, P. H., & Gluch, D. P. (2012). *Model-based engineering with AADL : an introduction to the SAE Architecture analysis and design language* (1st edition). Addison Wesley.
39. INCOSE Technical Operations. Systems engineering vision 2020. Version 2.03. Seattle, WA: International Council on Systems Engineering, INCOSE-TP-2004-004-02. Available Online: https://sdincose.org/wp-content/uploads/2011/12/SEVision2020_20071003_v2_03.pdf Accessed March 2025.
40. Trope, R. L. (2004). A warranty of cyberworthiness. *IEEE Security & Privacy*, 2(2):73–76
41. Boswell, L., Keane, J. & Mooney-Collett, C., (2017). Cyber test and evaluation in a maritime context. *Australian Defence Force Journal*, (202), pp.9–16.
42. van der Linden, D., & Rashid, A. (2022). The Effect of Software Warranties on Cybersecurity. *Software Engineering Notes*, *43*(4), 31–35. https://doi.org/10.1145/3282517.3302398
43. Schinas, O., & Metzger, D. (2023). Cyber-seaworthiness: A critical review of the literature. *Marine Policy*, *151*, 105592-. https://doi.org/10.1016/j.marpol.2023.105592
44. Dixon v. Sadler. 5 M. & W. 405. 1839.
45. McFadden v. Blue Star Line. 1 KB 697. 1905.
46. Marine Insurance Act, 1906, Marine Insurance Act 1906, section 39, https://www.legislation.gov.uk/ukpga/Edw7/6/41/section/39. accessed on 2023–02-23.

47. Australian Department of Defence (2024). Defence Seaworthiness Regulatory System Publication 100 – System Design and Requirements, Available Online: https://www.defence.gov.au/sites/default/files/2024-10/DEFENCESEAWORTHINESSREGULATORYSYSTEMPUBLICATION100.pdf Accessed March 2025
48. Suo, D., Siegel, J. E., & Sarma, S. E. (2018). Merging safety and cybersecurity analysis in product design. *IET Intelligent Transport Systems*, *12*(9), 1103–1109. https://doi.org/10.1049/iet-its.2018.5323
49. Macher, G., A. Much, A. Riel, R. Messnarz and C. Kreiner (2017). *Automotive SPICE, safety and cybersecurity integration*. 10489: 273-285.
50. Altawairqi, A. and M. Maarek (2017). *Attack modeling for system security analysis: (position paper)*. 10489: 81-86.
51. Friedberg, I., K. McLaughlin, P. Smith, D. Laverty and S. Sezer (2017). "STPA-SafeSec: Safety and security analysis for cyber-physical systems." *Journal of Information Security and Applications* 34(P2): 183-196.
52. Schmittner, C., Ma, Z., Puschner, P. (2016). *Limitation and improvement of STPA-sec for safety and security co-analysis*. In: Skavhaug, A., Guiochet, J., Schoitsch, E., Bitsch, F. (eds.) SAFECOMP 2016. LNCS, vol. 9923, pp. 195–209. Springer, Cham
53. Couce-Vieira, A., S. H. Houmb and D. Ríos-Insua (2018). *CSIRA: A method for analysing the risk of cybersecurity incidents*. 10744: 57-74.
54. Couretas, J. M. (2019). *An introduction to cyber modeling and simulation*. Hoboken, NJ, Hoboken, NJ : John Wiley & Sons, Inc.
55. Shahzad, S., Joiner, K., Qiao, L., Deane, F., & Plested, J. (2024). Cyber Resilience Limitations in Space Systems Design Process: Insights from Space Designers. *Systems (Basel)*, *12*(10), 434-. https://doi.org/10.3390/systems12100434
56. Smith, J. (2024). Ensuring Cyber Resiliency for OT Systems. *InTech*, *71*(2), 12–18.
57. Tjoa, S., Gafić, M., & Kieseberg, P. (2024). *Cyber Resilience Fundamentals* (1st ed. 2024.). Springer International Publishing. https://doi.org/10.1007/978-3-031-52064-8
58. International Organization for Standardization (ISO), I.: ISO/IEC 27001:2022: Information security, cybersecurity and privacy protection - information security management systems - requirements. Available Online: https://www.iso.org/standard/27001 (2022). Accessed March 2025
59. National Institute of Standards and Technology. Cybersecurity Framework. Available Online: https://www.nist.gov/cyberframework Accessed March 2025
60. Ross, R., Winstead, M., McEvilley, M.: NIST Special Publication 800-160, Volume 1: Engineering Trustworthy Secure Systems (2022). https://doi.org/10.6028/NIST.SP.800-160v1r1
61. Ross, R., Pillitteri, V., Graubart, R., Bodeau, D., Mcquaid, R.: NIST Special Publication 800-160, Volume 2: Developing Cyber-Resilient Systems: A Systems Security Engineering Approach (2021). https://doi.org/10.6028/NIST.SP.800-160v2r1
62. Carias, J. F., Arrizabalaga, S., Labaka, L., & Hernantes, J. (2021). Cyber Resilience Self-Assessment Tool (CR-SAT) for SMEs. *IEEE Access*, *9*, 80741–80762. https://doi.org/10.1109/ACCESS.2021.3085530

63. Kott, Alexander., & Linkov, Igor. (Eds.). (2019). *Cyber Resilience of Systems and Networks*. Springer International Publishing. https://doi.org/10.1007/978-3-319-77492-3
64. Segovia-Ferreira, M., Rubio-Hernan, J., Cavalli, A., & Garcia-Alfaro, J. (2024). A Survey on Cyber-Resilience Approaches for Cyber-Physical Systems. *ACM Computing Surveys*, *56*(8), 1–37. https://doi.org/10.1145/3652953
65. Shostack, A. (2014). *Threat modeling : designing for security* (1st edition). John Wiley and Sons.
66. Kordy, B., L. Piètre-Cambacédès and P. Schweitzer (2014). "DAG-based attack and defense modeling: Don't miss the forest for the attack trees." *Computer science review*. 13-14(C): 1-38.
67. Weiss, J.D. (1991). A system security engineering process. in: 14th Annual NCSC/NIST *National Computer Security Conference*: 572–581.
68. Schneier, B., Attack trees: modeling security threats, *Dobb's J. Softw. Tools* 24 (12) (1999) 21–29. Available at https://www.schneier.com/academic/archives/1999/12/attack_trees.html.
69. Paul, S. and R. Vignon-Davillier (2014). "Unifying traditional risk assessment approaches with attack trees." *Journal of Information Security and Applications* 19(3): 165-181.
70. Kordy, B., M. Pouly and P. Schweitzer (2016). "Probabilistic reasoning with graphical security models." *Information sciences* 342: 111-131.
71. J. Willemson, A. Jürgenson, Serial model for attack tree computations, in: D. Lee, S. Hong (Eds.), ICISC, in: *LNCS*, vol. 5984, Springer, 2010, pp. 118–128.
72. L. Wang, C. Yao, A. Singhal, S. Jajodia, Interactive analysis of attack graphs using relational queries, in: E. Damiani, P. Liu (Eds.), *Data and Applications Security* XX, in: LNCS, vol. 4127, Springer, Berlin Heidelberg, 2006, pp. 119–132. URL http://dx.doi.org/10.1007/11805588_9.
73. C. Phillips, L.P. Swiler, 1998. A graph-based system for network-vulnerability analysis. in: *Proceedings of the 1998 Workshop on New Security Paradigms* (NSPW'98). Charlottesville, Virginia, USA, pp. 71–79.
74. L.P. Swiler, C. Phillips, D. Ellis, S. Chakerian, 2001. Computer-attack graph generation tool. *DARPA Information Survivability Conference and Exposition* II (DISCEX'01) 2, 307–321.
75. L. Wen-ping, L. Wei-min, 2011. Space Based information system security risk evaluation based on improved attack trees. in: *Third International Conference on Multi-media Information Networking and Security* (MINES'11). pp. 480–483
76. Gribaudo, M., M. Iacono and S. Marrone (2015). "Exploiting Bayesian Networks for the analysis of combined Attack Trees." *Electronic Notes in Theoretical Computer Science* 310: 91-111.
77. Holm, H., M. Korman and M. Ekstedt (2015). "A Bayesian network model for likelihood estimations of acquirement of critical software vulnerabilities and exploits." *Information and software technology* 58: 304-318.
78. Moskowitz, I. and M. Kang (1998). *An insecurity flow model*. 129303: 61-74.
79. McDermott, J. and C. Fox (1999). *Using abuse case models for security requirements analysis*. 133431: 55-64.

80. Sindre, G. and A. L. Opdahl (2000). "Eliciting security requirements by misuse cases." *Proceedings of the Conference on Technology of Object-Oriented Languages and Systems*, TOOLS(2000): 120-131.
81. Karpati, P., G. Sindre and A. L. Opdahl (2010). *Visualizing cyber attacks with misuse case maps*. 6182: 262-275.
82. Abdulrazeg, A. A., N. M. Norwawi and N. Basir (2012). *Security metrics to improve misuse case model.* 94-99.
83. Saleh, F. and M. El-Attar (2015). A scientific evaluation of the misuse case diagrams visual syntax. *Inf. Softw. Technol*. 66: 73-96.
84. Mai, P., A. Goknil, L. Shar, F. Pastore, L. C. Briand and S. Shaame (2018). Modeling Security and Privacy Requirements: a Use Case-Driven Approach. *Inf. Softw. Technol*. 100: 165-182.
85. MatuleviÄius, R. (2017). *Fundamentals of Secure System Modelling*. Springer International Publishing. https://doi.org/10.1007/978-3-319-61717-6
86. Sindre, G. (2007). *Mal-activity diagrams for capturing attacks on business processes*. 4542: 355-366.
87. Opdahl, A. and G. Sindre (2009). Experimental comparison of attack trees and misuse cases for security threat identification. *Information and Software Technology* 51(5): 916.
88. Karpati, P., Y. Redda, A. Opdahl and G. Sindre (2014). Comparing attack trees and misuse cases in an industrial setting. *Information and Software Technology* 56(3): 294.
89. Tondel, I. A., J. Jensen and L. Rostad (2010). *Combining Misuse Cases with Attack Trees and Security Activity Models*: 438-445.
90. Meland, P. H., I. A. Tøndel and J. Jensen (2010). *Idea: Reusability of threat models - Two approaches with an experimental evaluation*. 5965: 114-122.
91. Department of Defense Instruction (DoDI) 5000.89 “Test and Evaluation”, November 19, 2020. Available Online: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/DoDI%205000.89%20Test%20and%20Evaluation.pdf Accessed March 2025
92. Booz Allen Hamilton Inc. (2023). Mission-based Cyber Risk Assessment. Available Online: https://www.boozallen.com/content/dam/home/docs/natsec/mbcra-slick-sheet.pdf Access March 2025
93. Sentient Digital Inc. Cyber Risk Assessment Tools and Risk Management: The MBCRA Process Available Online: https://sdi.ai/blog/cyber-risk-assessment-tools-and-risk-management-the-mbcra-process/ Accessed March 2025
94. AE Solutions (2023). Cyber Test and Assessment. Available Online: https://aesolutionshsv.com/cyber-test-%26-assessment Accessed March 2025
95. Defense Acquisition University. Mission Based Cyber Risk Assessment with USS Secure Available Online: https://www.dau.edu/blogs/mission-based-cyber-risk-assessment-uss-secure Accessed March 2025
96. Director Operational Test and Evaluation, US DoD. F-35 Joint Strike Fighter Report FY2024 Available Online:

https://www.dote.osd.mil/Portals/97/pub/reports/FY2024/dod/2024f-35jsf.pdf?ver=EUvCQMQSRdif89Gl9nye_g%3D%3D Accessed March 2025
97. Office of the Undersecretary of Defense, Research and Engineering, US DoD. Developmental Test, Evaluation and Assessment – Operational Resilience in Cyberspace Available Online: https://www.cto.mil/dtea/cyber/ Accessed March 2025
98. US Department of Defense (2020). Cybersecurity Test and Evaluation Guidebook, version 2.0. 10 February 2020. Available Online: Ohttps://ac.cto.mil/wp-content/uploads/2020/09/cyber-te-guide-v2c1.pdf Accessed March 2025
99. Kuzio de Naray, R; Buytendyk, A.M. Analysis of Mission Based Cyber Risk Assessments (MBCRAs) Usage in DoD's Cyber Test and Evaluation. Institute for Defense Analyses, *IDA Publication* P-33109, June 2022.
100. US DoD (2021). The Department of Defense Cyber Table Top Guide, Version 2, 16 September 2021. Available Online: https://www.cto.mil/wp-content/uploads/2023/06/DoD-Cyber-Table-Top-Guide-v2-2021.pdf Accessed March 2025
101. Monroe, M; Olinger, J. Mission-Based Risk Assessment Process for Cyber (MRAP-C). *The ITEA Journal of Test and Evaluation* 2020, Volume 41, pp. 229-232.
102. Jauhar, S., B. Chen, W. G. Temple, X. Dong, Z. Kalbarczyk, W. H. Sanders and D. M. Nicol (2015). Model-Based Cybersecurity Assessment with NESCOR Smart Grid Failure Scenarios. *Dependable Computing* (PRDC), 2015 IEEE 21st Pacific Rim International Symposium on: 319-324.
103. Nguyen P.H; Ali, S; Yue, T. Model-based security engineering for cyber-physical systems: A systematic mapping study. *Information and software technology*, 2017, Volume 83, pp. 116–135.
104. Geismann, J; Bodden, E. A systematic literature review of model-driven security engineering for cyber–physical systems. *The Journal of systems and software*, 2020, Volume 169
105. Carter, B; Adams, S; Bakirtzis, G; Sherburne, T; Beling, P; Horowitz, B. A preliminary design-phase security methodology for cyber–physical systems. *Systems* (Basel), 2019, Volume 7(2), pp.21.
106. Larsen, M.H; Muller, G; Kokkula, S. A Conceptual Model-Based Systems Engineering Method for Creating Secure Cyber-Physical Systems. *INCOSE International Symposium*. 2022, pp. 202–213.
107. Japs, S; Anacker, H; Dumitrescu, R. SAVE: Security & safety by model-based systems engineering on the example of automotive industry. *31st CIRP Design Conference*, 2021.
108. Bernardi, S., Gentile, U., Marrone, S., Merseguer, J., & Nardone, R. (2021). Security modelling and formal verification of survivability properties: Application to cyber–physical systems. *The Journal of Systems and Software*, 171, 110746-. https://doi.org/10.1016/j.jss.2020.110746

109. Navas, J; Voirin, J; Paul, S; Bonnet, S. Towards a model-based approach to systems and cybersecurity: Co-engineering in a product line context. *Insight* (International Council on Systems Engineering). 2020, Volume 23, pp. 39–43.
110. Geismann, J; Gerking, C; Bodden, E. Towards ensuring security by design in cyber-physical systems engineering processes. *International Conference on the Software and Systems Process*. 2018, Gothenburg, Sweden.
111. Mažeika, D; Butleris, R. MBSEsec: Model-based systems engineering method for creating secure systems. *Applied Sciences*. 2020, Volume 10(7), pp. 2574.
112. Object Management Group. (2022) UAF: Unified architecture framework. Available Online: https://www.omg.org/spec/UAF Accessed March 2025.
113. Jurjens, J. *Secure Systems Development with UML*. Berlin, Heidelberg: Springer; 2005.
114. Apvrille, L; Roudier, Y. Towards the model-driven engineering of secure yet safe embedded systems. *International Workshop on Graphical Models for Security*. 2014.
115. Dassault Systemes (2025) CATIA Magic: Global Model-Based Systems Engineering Solutions. Available Online: https://www.3ds.com/products/catia/catia-magic Accessed March 2025
116. Object Management Group. (2017) OMG Systems Modeling Language (SysML) version 1.5. Available Online: https://www.omg.org/spec/SysML/1.5/ Accessed March 2025.
117. Fowler, S.; Sitnikova, E. Toward a framework for assessing the cyber-worthiness of complex mission critical systems. In *Proceedings of the 2019 Military Communications and Information Systems Conference* (MilCIS), Canberra, Australia, 12-14 November 2019.
118. Fowler, S. Cyber Evaluation and Management Toolkit (CEMT). Available Online: https://github.com/stuartfowler/CEMT/ Accessed March 2025
119. Allen, M. S., Robson, D. A., & Iliescu, D. (2023). Face Validity: A Critical but Ignored Component of Scale Construction in Psychological Assessment. *European Journal of Psychological Assessment : Official Organ of the European Association of Psychological Assessment*, 39(3), 153–156. https://doi.org/10.1027/1015-5759/a000777
120. Royal, K. (2016). 'Face Validity' is Not a Legitimate Type of Validity Evidence. *The American Journal of Surgery*, 212(5), 1026–1027. https://doi.org/10.1016/j.amjsurg.2016.02.018
121. MITRE. ATT&CK Matrix for Enterprise. Available Online: https://attack.mitre.org/ Accessed March 2025