Assessing Cyber-Worthiness of Complex System Capabilities using MBSE: A new rigorous engineering methodology

Stuart Fowler*, Keith Joiner[†] and Elena Sitnikova[‡]

UNSW Canberra, University of New South Wales
Canberra, Australia

Email: *stuart.fowler@student.unsw.edu.au, [†]k.joiner@unsw.edu.au, [‡]e.sitnikova@unsw.edu.au

Abstract—Cyber-worthiness as it is termed in Australian Defence, or cyber-maturity more broadly, is a necessary feature of modern complex systems which are required to operate in a hostile cyber environment. To evaluate the cyber-worthiness of complex systems, an assessment methodology is required to examine a complex system's or system-of-system's vulnerability to and risk of cyber-attacks that can compromise such systems. This assessment methodology should address the cyber-attack surface and threat kill chains, including supply chains and supporting infrastructure. A cyber-worthiness capability assessment methodology has been developed based on model-based systems engineering concepts to analyse the cyber-worthiness of complex systems and present a risk assessment of various cyber threats to the complex system. This methodology incorporates modelling and simulation methods that provide organisations greater visibility and consistency across diverse systems, especially to drive cybersecurity controls, investment and operational decisions involving aggregated systems. In this paper, the developed methodology will be presented in detail and hypothesised outcomes will be discussed.

I. INTRODUCTION

Our introduction provides a high-level overview of what has been developed and why since an earlier published paper [1]. Please see the Background section and references for our critical research influences and explanations of any key new terms.

This paper details a methodology for assessing the cyber-worthiness or cyber-maturity of complex systems like critical infrastructure and weapon systems using model-based systems engineering (MBSE). The basis for this methodology is the development of a detailed cyber-attack surface representation and threat model using SysML, the modelling of architectural and security controls to mitigate the threats identified in the threat model, and then the simulation of the model for a sensitivity analysis that assesses risk against each identified threat and overall. These three phases of the methodology can be performed sequentially or iteratively in the cyber-worthiness assessment and are self-contained, allowing for the methodology to be tailored by the modeller to suit different systems and their lifecycles. The framework was developed primarily for complex systems or capabilities that of themselves are not information communication technology (ICT) capabilities but since the Information Age and with increasing software functionality, all contain a degree of

embedded information technology. Examples include ships, aircraft, hospital theatres and transportation systems. Such engineered systems are arguably more vulnerable today than dedicated ICT systems because of the following factors:

- they are usually updated less regularly [2],
- they are often more bespoke and mix proprietary architectures with open-source elements [2],
- are usually less well monitored as aspects are still trusted (designed and tested for use and abuse but not malicious intents) [3],
- because of the need to partner specialty platform engineers with ICT cybersecurity engineers in any assessment [4] [5],
- because of the need to link and instrument specialty representative platform sites (hardware-in-the-loop and software integration facilities/equipment) with cyber-ranges [6] [7],
- are targeted by advanced Nation states [4] [8], and
- if successfully compromised a Nation-state cyber-agent is likely to store a procedure against critical infrastructure or weapon systems for later coordinated use (i.e., cyber-storming) with as little evidence of compromise as possible [9] [10].

A. Background

Cyber-worthiness is a recent Australian Defence Department term created based on a history of a type of technical regulation in a Westminster or Commonwealth country where Defence has an obligation for internal regulation for safety and its own use of air, sea and land spaces. The more recent history began in the 1990s with airworthiness regulations and governance [11] [12] and this flowed through to seaworthiness [13] and land-worthiness [14]. The 'worthiness' governance has a degree of regulation and monitored assurance over design, manufacture, resupply, engineering, and maintenance. As such 'worthiness' is synonymous with initial and continued safety and functionality, dictating the approved organisations, people, processes and data. Hence, these regulations are applied more rigorously based on safety criticality. The assurance concepts are synonymous with general governance maturity models, such as program management offices, however, are arguably

more controlling, intrusive and onerous than civilian equivalents. Cybersecurity has challenged Australian Defence, as it has U.S. Defense, though arguably still lagging in processes in Australia by many years [4] [7] [15] [16]. As cybersecurity threats became more prevalent and necessary across more safety-critical systems that were not primarily ICT, it became necessary to apply broader and more enforceable assurances than ICT acquisition strategies, hence cyber-worthiness was first termed [3] [17] and is defined as [18]:

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.

The concept is analogous to the U.S. term cybersecurity maturity, especially the recent U.S. Defense Cybersecurity Maturity Model Certification (CMMC) [19] where the concepts map well, especially the certification processes to regulatory assurance. However, these governance concepts will not succeed without systematic engineering frameworks for assessment. According to Easttom and Butler [20]:

It is unfortunately the case that penetration testing is frequently conducted in an ad hoc, imprecise manner (Easttom, 2018). There are standards to guide penetration testing, but these are often generalized guidelines, without specificity regarding techniques. What is currently absent from the current literature is a defined, engineering approach to penetration testing.

Easttom and Butler [20] then recommend systems modelling with misuse cases to properly prepare for penetration testing. The cyber-worthiness assessment methodology outlined in this paper therefore seeks to address this concern directly. In a separate paper, Easttom [21] addresses one of the other main deficiencies in cybersecurity, that of cyber threat modelling:

There currently exists a wide array of methods to examine cyber-attacks. Many of those are effective in specific narrow scenarios. These methods allow for analysis and improved understanding of cyber-attacks, but they have limited applicability. What is absent from the literature are two elements. The first element is a broadly applicable method to model and analyse any cyber-attack. The second issue that is absent, is the precise mathematical quantifying of attacks.

Easttom [21] suggests using several failure analysis techniques, such as graph theory. Our cyber-worthiness assessment framework seeks to address the threat modelling concerns, however, uses bow-tie diagrams from safety theory and traditional cyber-attack trees to perform the failure analysis. The quantifying of attacks remains difficult to estimate and will be discussed further in limitations and future research. While the cyber-worthiness assessment framework does account for these estimates, perhaps through consistent use and modelling and improved awareness and recording of threat incidence,

risk assessments can be evolved with current threat data to achieve mathematical quantification.

In summary, both the Australian cyber-worthiness and U.S. CMMC governance issues warrant consistent assessment frameworks that leverage more rigorous and consistent engineering approaches to prepare for and use precious real penetration testing to best effect. The assessment framework developed aims to provide that rigorous approach.

B. Overview

Each section of this paper provides an example of the diagrams used to generate the model underpinning the cyber-worthiness assessment methodology. Our Cyber-Worthiness Evaluation and Management Toolset (CEMT) has been developed by extending the built-in SysML classes, and the taxonomy used in this toolset is described in this paper to show an example implementation of the underlying methodology. In general, the toolset uses *Stereotypes* to identify the various objects used in the methodology, rather than using the base SysML classes directly. This helps to ensure separation between the modelling activities used as part of the cyber-worthiness assessment methodology, which helps to reduce conflicts in model queries when integrating the cyber-worthiness assessment methodology into a broader SysML model. The CEMT provides a *Profile of Stereotypes and Customizations* that can be used in any SysML project, which aids in re-usability and ensures a consistent set of queries between systems.

Fig. 1. Steps in the Cyber-worthiness Assessment Methodology

The paper follows the three primary processes within the cyber-worthiness assessment methodology, which are shown in Figure 1. Threat Modelling encompasses the activities associated with developing the threat model that describes the attack surface and threat kill chains. Threat Mitigation refers to the identification of potential security controls and mitigations to prevent or impede malicious threat actors. Cyber-worthiness Assessment involves the development of a transparent and traceable risk assessment that can be used as the basis for a cyber-worthiness claim.

II. THREAT MODELLING

The threat modelling phase is a mechanism for identifying and documenting the threats to the complex system. The threat model uses the concepts of Misuse Cases and Mal-Activity diagrams to develop a threat model by drawing a Directed Acyclic Graph (DAG) of the steps that an adversary needs to take in order to compromise the system. The primary benefits of documenting this threat model in SysML are the ability to query the threat model to develop specific views of the model and the ability to integrate the threat model into other

Fig. 2. Stereotype Definitions for Misuse Case Diagrams in the Cyber-worthiness Evaluation and Management Toolset (CEMT)

Fig. 3. Misuse Case Diagram in the Cyber-worthiness Evaluation and Management Toolset (CEMT)

system design artefacts that leverage model-based systems engineering techniques.

A. Misuse Cases

Misuse Cases are modelled using a standard *SysML Use Case Diagram* with the stereotypes shown in Figure 2. The *CyberActor* stereotype was created with the *Actor* metaclass, and this is used as the base classifier for two further stereotypes - *MaliciousActor* and *NonMaliciousActor*. A *MisuseCase* stereotype is also created with *UseCase* as the metaclass. These stereotypes are added to a *SysML Use Case Diagram* and the actors are associated with the misuse cases in which they participate, as shown in Figure 3.

B. Mal-Activity Diagrams

Mal-Activity Diagrams are modelled using the standard *SysML Activity Diagrams* with the stereotypes shown in Figure 4. Most of these stereotypes are based on the standard objects used to produce a *SysML Activity Diagram*, and most are

Fig. 4. Stereotype Definitions for Mal-Activity Diagrams in the Cyber-worthiness Evaluation and Management Toolset (CEMT)

simple stereotypes used to distinguish those objects used as part of the Cyber-Worthiness Evaluation and Management Toolset (CEMT) from other objects used in activity diagrams in other parts of the system model. However, some have more complicated properties, and those are expanding on below.

ThreatModelFlow is based on the *ControlFlow* metaclass, and is then used as the base classifier for both *ThreatFlow* and *DetectionFlow*, which are used to create the flows between the actions of the threat actor and the actions of the system's detection efforts, respectively.

ThreatModelAction is based on the *CallBehaviorAction* metaclass, and is used as the base classifier for *DetectionAction* and *ThreatAction* which are the primary nodes used in the mal-activity diagrams. *ThreatAction* is used to represent the actions of a threat actor in the mal-activity diagram, and *DetectionAction* is used to represent the actions of either the system being attacked, or the operators defending against the threat actor. The *AggregatedAction* stereotype is used as an intermediate node which has further threat model nodes nested inside.

The *ThreatAction* stereotype is given a *Difficulty* attribute, which is used to capture how difficult a particular threat action is for a malicious actor to complete. As a broad example, the use of valid credentials to log into a system may have a trivial difficulty, while a side-channel attack that looked at electromagnetic emissions from a processor would have a higher difficulty level.

ThreatAction is further modified using a *Customization*, which adds three *Derived Attributes* to the stereotype. The first attribute is NextThreatAction which uses a series of queries to traverse the ThreatFlows leaving the node to identify which nodes are connected as the next step in the mal-activity diagram. The second attribute - PreviousThreatAction - does the same thing in the opposite direction to find the preceding step(s) in the mal-activity diagram. The final derived property is DetectionAction which traverses the DetectionFlows leaving the node to determine what actions to detect the malicious action have been modelled in the mal-activity diagrams.

ThreatModelAction also has a *Customization* applied which renames the *AllocatedTo* attribute to allocatedComponent and the *AllocatedFrom* attribute to potentialControl. These attributes are used to link the components that are relevant to the ThreatModelAction and to security controls that may be used to mitigate the particular action, respectively. Due to inheritance both ThreatAction and DetectionAction have this customization applied to them.

The ThreatSendSignal and ThreatAcceptEvent stereotypes, which inherit from the *SendSignalAction* and *AcceptEventAction* metaclasses respectively, are customized to include the LinkedDiagram derived property. These stereotypes are used to generate signals and events that allow different activity diagrams to be linked together. The LinkedDiagram property uses chained queries to traverse those signals, and list the other activity diagrams that these objects can trigger or can be triggered from.

These *Customizations* provide a set of derived properties that defines convenient variables that show the sequential structure of the directed graph that is being modelled in the various mal-activity diagrams. These are used in later steps of the methodology to automatically create various views of the threat model structure to aid in threat mitigation and assessment.

Figure 5 provides an example of a mal-activity diagram for the Insider Threat misuse case. In this example, the threat begins at the *InitialNode* on the left and progresses along the ThreatFlows, through the various nodes of the model and finishing at one of the ThreatImpacts, namely a Confidentiality Loss, an Integrity Loss or an Availability Loss. The system detection thread is also modelled, with each threat action having a chance to be detected, culminating in the ThreatDetection called Malicious Activity Detected.

This diagram shows the use of ThreatAcceptEvents - the System Access and Room Access nodes - which accept triggering signals from other mal-activity diagrams. These allow for the threat model sequences modelled as part of this Insider Threat misuse case to be reused by the mal-activity diagrams associated with other misuse cases.

Figure 5 also shows the use of AggregatedNodes to show the nesting of more complicated detail in the threat model. Figure 6 shows the mal-activity diagram tied to the *Access System AggregatedAction*.

While AggregatedActions can contain more AggregatedActions for multiple levels of nesting, at the lowest level of the nesting, as shown in Figure 6, the mal-activity diagram consists of just ThreatActions, DetectionActions and the control flows between them.

Note that the ThreatActions are assigned Difficulties and guard conditions are used to show that the threat model only proceeds if the particular action was successful. The [else] keyword is used to show that when a step fails, this particular threat thread ends. Additionally, there is no guard condition on the DetectionFlow between the ThreatAction and DetectionAction. This is because that Component Activity is always generated when the threat action is attempted, giving the system the ability to detect that event. The DetectionAction itself, however, does have guard conditions on its outputs, as there is a probability that the detection will fail.

III. THREAT MITIGATION

The Threat Mitigation phase of the cyber-worthiness assessment methodology takes the mal-activity diagrams developed in the Threat Modelling phase, and attaches mitigation techniques to each of the nodes of the mal-activity diagrams. The stereotypes shown in Figure 7 are used in order to model these mitigations. The Asset stereotype is a simple object which is based on the *Block* metaclass, and is used to tag those components within the system that are relevant to the threat model. This could be a stereotype added to the *Blocks* used in an existing system model, or new blocks specifically created for the threat model, as determined by the modeller.

The SecurityControl stereotype inherits from the *Requirement* metaclass, and is used to model potential controls and mitigation techniques which could be used to either prevent a malicious actor from performing one or more of the actions in the mal-activity diagrams or contribute to one or more of the detection actions in the mal-activity diagrams. This could be a stereotype added to the *Requirements* used in an existing system model, or new controls specific for the threat model.

Figure 8 is an extension of the mal-activity diagram shown in Figure 6, and shows how Assets and SecurityControls are linked into the threat model. The Assets are linked into the ThreatAction and DetectionAction nodes using the allocatedComponent relationship - which is simply a renaming of the default *AllocatedTo* relationship - to show which components in the system are related to a particular node in the threat model. The SecurityControls are also linked into the ThreatAction and DetectionAction, but they use the potentialControl relationship - which is simply a renaming of the default *AllocatedFrom* relationship. These security controls describe the potential controls that could be implemented by the system in order to mitigate (in the case of a malicious action) or perform (in the case of a detection action) the specific node in the threat model.

Fig. 5. Top level mal-activity diagram for the *Insider Threat* misuse case

Fig. 6. Nested mal-activity diagram for the *Access System* AggregatedAction

It is important to note that this isn't just a listing of the currently implemented controls, or controls which are planned to be implemented. The security controls linked to each node of the threat model should represent a comprehensive set of all security controls that could feasibly be implemented to address that specific node. This approach during the Threat Mitigation phases enables the development of So Far As Reasonably Practicable (SFARP) arguments in the Cyber-worthiness Assessment phase.

The SecurityConstraint stereotype, also shown in Figure 7, is derived from the *Property* metaclass and contains an *Implementation* attribute. The SecurityConstraint stereotype is used to model the instantiation of a SecurityControl as a property of an Asset, and the Implementation attribute is used to identify whether that particular SecurityControl has been implemented on a specific Asset. With reference to Figure 8, the SecurityConstraints are generated

Fig. 7. Stereotype Definitions for Threat Mitigation in the Cyber-worthiness Evaluation and Management Toolset (CEMT)

by taking each of the Assets allocated to a threat model node and creating a property within that Asset object for each of the SecurityControls linked to the same threat model node. Those new properties are given the SecurityConstraint stereotype, and the SecurityControl from which they were derived is used as the *Type* of the new SecurityConstraint.

The end result can be seen in Figure 9, where the Asset hierarchy can be built with the SecurityConstraints captured as properties of each Asset. As this Threat Mitiga-

Fig. 8. Nested mal-activity diagram for the *Access System* AggregatedAction with linked *SecurityControls* and *Assets*

Fig. 9. *SecurityConstraints* as properties of *Assets* in the Cyber-worthiness Evaluation and Management Toolset (CEMT)

tion phase is performed on each of the mal-activity diagrams underneath the misuse cases, this becomes a list of potential mitigation techniques for each component in the system, derived directly from a threat model and traceable back to the specific malicious action(s) and/or detection opportunities that warrant the inclusion of that specific mitigation technique. In essence, this creates a set of potential security requirements that are inherently contextualised to the system and the threat environment in which it operates.

The final step in the Threat Mitigation phase is to audit the implementation state of each of the identified *SecurityConstraints* and update the *Implementation* attribute of each object accordingly. The ability to accurately determine this implementation state is dependent on the maturity of the system. If the system is still within the early design phases, this might be a planned implementation state and the follow-on Cyber-worthiness Assessment phases can be used to validate whether that planned implementation state will align with the risk appetite of the system owners. If the system is already operational and this methodology is used as part of a cyber table topping activity, or a cyber test and evaluation activity, then

the implementation state of each control may be directly auditable.

IV. CYBER-WORTHINESS ASSESSMENT

The Cyber-worthiness Assessment phase involves the examination and critical review of the threat model by stakeholders in order to understand and evaluate the cyber-worthiness of the system. While it is difficult to eliminate subjectivity and expert judgement in this sort of evaluation, this phase aims to remove or reduce opaque expert judgement and unsupported subjectivity. The traceability and granularity modelled during the earlier phases of the methodology allow the modellers to underpin their decisions with documented rationale and consistency.

The Cyber-worthiness Assessment phase can be segmented into three primary activities.

- 1) Summary Diagrams - which provide consumable and intuitive views of the threat model;
- 2) Risk Assessment - which articulates the risks associated with each threat; and
- 3) Simulation - which performs a numerical simulation of the threat to determine risk probabilities.

This phase is distinctly different from the previous phases, as this is the step where non-experts get directly involved in the threat model. The focus of the methodology shifts from modelling the specific cyber threats toward the presentation of the data in a manner that is easily consumable and engaging for non-experts.

A. Summary Diagrams

Summary diagrams are relationship maps that are automatically generated based on the threat modelling and threat mitigation activities performed in the earlier phases. These diagrams are used to take the specific detail of the nested threat model and present it in a manner that is intuitive and explainable to non-experts. This is critically important, as the designers and owners of complex systems are often not cybersecurity subject matter experts, but they are the ones that are asked to accept the residual risk associated with any cybersecurity risk. Presenting the information in a manner that allows those people to make informed decisions is a critical component of the cyber-worthiness assessment methodology.

The first type of summary diagram is the Attack Tree, a partial example of which is shown in Figure 10. This diagram leverages the *NextThreatAction* and *PreviousThreatAction* derived properties that were added to the stereotypes of each node in the mal-activity diagram in order to produce a tree of the threat steps that must be taken by a malicious actor in order to achieve a loss of Confidentiality, Integrity or Availability. While the example shown in Figure 10 starts at a particular misuse case (Insider Threat) and will show all of the paths that could be taken from that starting point, the reverse of that can also be generated in order to show a particular outcome (eg. Confidentiality Loss) and all the possible paths leading to that outcomes from all misuse cases.

Fig. 10. Partial attack tree for the *Insider Threat* misuse case in the Cyber-worthiness Evaluation and Management Toolset (CEMT)

Fig. 11. Partial bow-tie diagram for the *Insider Threat* misuse case in the Cyber-worthiness Evaluation and Management Toolset (CEMT)

The attack tree diagram provides a high level overview of the threat model. It can be used by modellers and stakeholders to identify the full kill chain associated with a particular malicious attack, and facilitates the identification of key activities that may be a common node between many attacks. This can highlight the key activities that can become of the focus of any additional mitigation efforts. Similarly, by explicitly enumerating every step that is required in order to achieve a particular malicious outcome, defence-in-depth principles can be readily implemented by ensuring mitigations are applied at as many steps in the chain as possible.

The second type of summary diagram is the bow-tie diagram, shown in Figure 11. This provides a view of the nested hierarchy of mal-activities and a summary of the `SecurityConstraints` that are associated with each threat node. The implementation state of those `SecurityConstraints` is also visible in the bow-tie diagram.

This diagram is inspired by the bow-tie diagrams commonly found in system safety engineering analyses, and is used as a way to provide a high level overview of how many potential controls have been implemented for a particular misuse case. It also provides a backdrop upon which SFARP-like discussions can take place. The bow-tie diagram not only shows what mitigating controls have been implemented, but also what further mitigations could feasibly be implemented, catalysing the discussion over whether those additional controls are worth implementing or not.

Of course, this is just two of the potential summary diagrams that can be developed from the threat model. Once the model is developed, the objects and relationships can be queried in order to rapidly develop additional representations of the modelling information.

A simple example would be a table showing all mitigating controls currently implemented on a particular `Asset`. This could become useful for an engineering support agency that was tasked with replacing that particular component, as it would provide a clear and concise list of the specific security controls that are contributing to the cyber-worthiness of

the system; as well as any additional controls which would improve the security posture.

Equally, a table showing the `SecurityControls` in the threat model and a count of how many threat model nodes they are linked to would provide an indication of which mitigations are most impactful and therefore are worth investigating more closely.

B. Risk Assessment

The risk assessment is one of the primary outputs of the threat model. It provides a measure of the risk associated with each threat in terms of a likelihood of a particular attack succeeding and the consequence of that attack. The stereotypes shown in Figure 12 are used to build the risk assessment model as part of the Cyber-worthiness Assessment phase.

The primary stereotype is `SecurityRisk`, which captures all of the relevant risk information for a particular kill chain. Attributes for Likelihood, Consequence and Risk Rating are included, along with free-text fields to provide a justification for the ratings. Also included are attributes to capture the results of any risk simulations, which will be discussed further in later sections of this paper.

The other stereotypes shown in Figure 12 are simple stereotypes used to create specific tags for the properties used in the risk assessment diagrams. These are used in the creation of the risk diagrams to ensure any queries of the model can easily select for these particular properties.

Figure 15 shows an example risk diagram for one particular kill chain. This is a SysML parametric diagram, which shows an Initial Probability in the top left being modified by a number of constraint blocks in order to determine a Residual Probability and a Detection Probability. Note that the structure of this parametric diagram is mirroring the sequence of actions along a single branch of the attack tree, with the relevant `DetectionActions` included. This means that these diagrams can be generated through scripts that query the model and then draw this diagram because there are no new relationships that must be defined in order to produce this diagram.

Fig. 12. Stereotype Definitions for Risk Assessments in the Cyber-worthiness Evaluation and Management Toolset (CEMT)

However, there are three exogenous variables that need to be added to the diagram in order to complete the Risk Assessment diagram:

- 1) Initial Probability;
- 2) Threat Level; and
- 3) Control Effectiveness

The `InitialProbability` variable is defined as the probability that the particular attack will be attempted within the period of interested. That period may be a certainty if the desired outcome is to get an engineering assessment of how well the system performs in response to an attack, or it could be a smaller percentage if the risk assessment is being performed for a specific mission or scenario.

The `ThreatLevel` variable is used to reflect the skill level of the threat actor, from a Novice up to a Nation State. The more capable the threat actor, the more likely they are to achieve a malicious action of a specific difficulty. This variable can be modified to assess the specific threat actor that is most relevant to the system's threat context.

`ControlEffectiveness` is defined as the effectiveness of the mitigating controls in preventing a particular threat action from occurring. More precisely, it is the percentage of attempted malicious actions that are thwarted by the mitigation controls. This is a value that is difficult to measure or even estimate precisely, and ultimately it is determined through expert judgement of the cybersecurity subject matter experts that are involved in the modelling process.

It is important to understand that these exogenous variables are not determined objectively. They are subjectively derived

from the judgement of subject matter experts. This is not unusual for generalised cyber risk assessment methodologies - the risk levels themselves are often determined through expert judgement. The benefit of the proposed methodology is that it attempts to remove *opaque* expert judgement. Making those decisions at the level of how well a particular control set mitigates a specific action provides much clearer traceability that making those same judgements at the level of a whole kill chain, or indeed an entire misuse case.

The risk assessment is often summarised in a table, which queries the model to generate a table of each assessed threat chain and outlines the malicious actions required that must be completed for the threat to succeed, a listing of the mitigating security controls and an indication of the further mitigating controls that could be added.

C. Simulation

While the risk assessment diagrams provide an overview of a specific kill chain and the different attributes of the system that act to mitigate or detect that kill chain, a mathematical simulation is required in order to calculate risk values. The simulation of the parametric diagram also allows for uncertainty in the exogenous variables that are determined through expert judgement and can also facilitate sensitivity analysis on those variables. This uncertainty can be modelled as a uniform distribution, as shown by the `min` and `max` values on the variables in Figure 15, or as a normal distribution with a mean and standard deviation.

Fig. 13. Simulated results of threat analysis in the Cyber-worthiness Evaluation and Management Toolset (CEMT)

Running a numeric simulation on these input distributions produces histograms for the `ResidualProbability` and `DetectionProbability` as shown in Figure 13 and Figure 14, respectively. These values can be attached to the `SecurityRisk` block associated with the kill chain as the `Simulation Residual Probability` and `Simulation Detection Probability` attribute and can then be used as supporting justification for allocating a particular qualitative likelihood, consequence and risk rating in accordance with the particular risk management framework being used. These simulated values, and the qualitative risk ratings are added to the risk summary tables produced as part of the earlier Risk Assessments in order to finalise the risk table for the system.

Fig. 14. Simulated results of detection analysis in the Cyber-worthiness Evaluation and Management Toolset (CEMT)

The ability to rapidly simulate the risk assessment facilitates two further activities:

- 1) Mission-based Reassessment; and
- 2) Contextualised SFARP discussions.

Mission-based reassessment allows for the risk to be determined on a mission-by-mission basis, by varying the initial probability of a particular threat based on specific threat intelligence, and by altering the implementation state of various mitigating controls to satisfy the new residual risk appetite. This can be useful in determining the new level of residual risk in a specific environment, or perhaps even to reassess the risk if a particular mitigating control is non-functional or otherwise not working as expected.

Contextualised SFARP discussions refers to the ability to make informed decisions on whether enough mitigating or detecting controls have been implemented, or whether more is required in order to satisfy the risk appetite of the system owner. The implementation status can be modified to simulate the inclusion of additional controls, and the control effectiveness updated accordingly. This would produce a new level of residual risk for that specific kill chain. This will provide a magnitude of residual risk reduction which can be compared with that achieved by other potential controls in order to make a judgement on whether the risk has been reduced so far as reasonably practicable. This provides a level of granularity in the residual risk assessment that is lost in many mandated risk management frameworks. Qualitative likelihood measures at the threat level are rarely moved by a single new mitigating control, but the simulated residual risk can show the percentage change in the residual risk level for each new control. This facilitates more informed and nuanced discussion of the costs and benefits of a new control.

V. DISCUSSION AND FUTURE DIRECTIONS

The methodology developed offers significant rigour to mature and broaden cybersecurity for cyber-worthy complex systems as called for by [21], especially where vehicle and industrial engineering have significant ICT technologies embedded but they are not the *raison d'être* of the capability. The techniques used are common in safety assessments and so industry has the demonstrated ability to apply these methods

where necessary. Wide usage offers the prospect of raising the ‘barrier to entry’ of cyber threats away from some of the criminal and untrustworthy states, thereby improving the prospects of attribution when attacks do still occur [10]. Constituent parts of the methodology have been applied in cyber-worthiness assessments of some major maritime capabilities; however, these were done as a ‘means to an end’ and without time to evaluate and improve the process. This research documents the overall method fully and will now progress to a rigorous two-year validation wherever the method can be applied. To assist that validation and given the scourge the method seeks to combat, wide dissemination before validation is apropos.

Greater consistency between system program offices and projects is enabled by implementation of the Cyber-worthiness Evaluation and Management Toolset (CEMT) and the nested mal-activity diagrams that allow for the threat model to be decomposed and then allocated to specific subprojects and subsystem designers. That nested detail can be pulled back into a consistent system-of-systems level threat model to allow the assessment of emergent properties in the system level analysis. Greater re-use of common modules and sub-systems is enabled by the modularity of the MBSE threat modelling approach which allows for the reusability of sections of the threat model. This reuse could be within a single threat model (i.e., reusing the post-exploit behaviours between the different threat vectors) or across models (i.e., reuse of the mal-activity diagrams for a component that is common between two systems).

More detailed, informed and nuanced cyber table-topping or risk assessments of capability during development or through-life decision-making is enabled primarily by the step-by-step modelling of malicious activity. This sequencing allows for non-subject matter experts to engage meaningfully in cyber table topping exercises. Abstract threat descriptions are replaced by explicit actions and the exogenous variables and the expert judgement behind the determination of those variables is clearly identified and can be interrogated. This minimises opaque expert judgement and allows for more informed decisions to be made by more engaged decision makers. Furthermore, the MBSE approach clearly identifies mitigations and controls that can work across several threat vectors. This allows for the relative impact of security controls to be determined to better inform investment decisions. Controls are also tied to specific assets and components, providing a clear list of contextually relevant security controls provided by a particular component to assist in obsolescence analysis.

The methodology aims to improve identification of the most important cases to verify the model with actual penetration testing by simulating in table-top risk exercises the multitude of threat vectors that may interact with the cyber-attack surface of a given system. The table-top exercise can then shortlist any vulnerability and penetration testing to ensure that resources and effort are focused on those areas that are most relevant to the system.

Most importantly for larger departments, the method enables greater efficiency and flexibility to aggregate cyber-worthiness

assessments on systems-of-systems because the MBSE approach allows for modifications to a threat model to be made with the confidence that the model remains internally consistent. This modelling language allows for modification to the threat model to be made rapidly to support ad hoc systems-of-systems that may be formed. Such an operational assessment could be as simple as modifying the initial likelihood of a particular mis-use case for a particular mission or as complex as extending the model to include the new components added into the system-of-systems. In either case, these modifications can be made in one place of the model, and all of the summary diagrams and risk assessments can be updated in near real time to determine the new risk levels in the new environment. Such adaptability, however, has its greatest potential benefit in through-life update to account for advanced persistent threats. The technical debt of maintaining complex attack trees, risk assessments and control implementation details is minimised as the MBSE approach allows for changes in one place to automatically update the other views of the model. Traceability of the controls provides clear justification for controls implemented in a system, and this can be re-evaluated and challenged when system refreshes, updates and upgrades introduce new features and capabilities which may render those controls less critical than they were previously.

The creation of a cyber-worthiness assessment framework based on MBSE is expected to deliver the following benefits.

- Greater consistency between system program offices and projects developing and sustaining critical infrastructure and weapon systems.
- Greater re-use of common modules and sub-systems between system program offices and projects within, and even between, departments.
- More detailed, informed and nuanced cyber table-topping or risk assessments of capability during development or through-life decision-making.
- More informed and thus fairer investment decisions at development iterations, technical refreshes or upgrade opportunities particularly posturing between investing in enhanced capability like greater enterprise information management compared to more resilient security controls.
- A more robust basis for determining the most important cases to verify the model with actual penetration testing, whereby the very many permutations of cyber-attack surface configurations, defensive posturing of controls, and threats, can be more easily analysed for representative sampling and combinatorial test design methods like high throughput testing.
- Greater efficiency and flexibility to aggregate cyber-worthiness assessments on systems-of-systems that are formed to meet operational demands, like deploying elements of a hospital to an emergency, forming transport for a one-off large event, or in military joint forces.
- Ease in updating cyber-worthiness assessments to new threats, sub-systems, suppliers or software refreshes. For Governments and militaries this likely advantage is in

keeping pace with advanced persistent threats (APTs).

These form the basis of our research questions for the next phase of investigation, where public service departments and their supporting industry are invited to adopt the framework and be interviewed on the strengths and weaknesses. The CEMT will be implemented on a number of complex systems in order to assess the effectiveness, efficiency and user satisfaction of the proposed methodology. This validation of the Cyber-worthiness Assessment Methodology will form the basis of subsequent papers.

VI. CONCLUSION

This paper outlined a new rigorous engineering methodology for the assessment of cyber-worthiness of complex systems that encompasses Threat Modelling, Threat Mitigation and a Cyber-worthiness Assessment. The Cyber-worthiness Evaluation and Management Toolset (CEMT), an MBSE meta-model that allows for the mindful, collaborative, accountable and transparent assessment, is presented as an implementation of this new methodology. The methodology provides an approach that stakeholders without deep cybersecurity knowledge to be engaged in the cyber risk assessment process, and aim to reduce or remove the opaqueness that is often associated with these sorts of risk assessments. By introducing rigour into this assessment process, the methodology is able to instill a level of assurance in the process which can be used to make claims related to the cyber-worthiness of a given system.

The use of a common toolset, such as CEMT, when implementing the cyber-worthiness assessment methodology also allows for the aggregation and integration of related systems into a system-of-systems to facilitate holistic risk assessments. By focusing on the interfaces between the threat models of the individual systems, the likelihood output from one model can be used as an input likelihood to the other model, and vice versa. This paves the way for capability level cyber-worthiness assessments that can be rapidly updated as the system-of-systems is changed. The precise definition of a common set of interfaces is a task for future research.

NOTES

The methodology has no intellectual property overlays as it is based on the sound adoption of engineering principles. Agencies seeking to participate in the validation activity should contact the corresponding author.

REFERENCES

- [1] S. Fowler and E. Sitnikova, "Toward a framework for assessing the cyber-worthiness of complex mission critical systems," in *Military Communications and Information Systems Conference*, Conference Proceedings.
- [2] G. Brennan, K. F. Joiner, and E. Sitnikova, "Architectural choices for cyber resilience," *Australian Journal of Multidisciplinary Engineering*, vol. 15, pp. 68–74, 2019.
- [3] S. Fowler, C. Sweetman, S. Ravindran, K. F. Joiner, and E. Sitnikova, "Developing cyber-security policies that penetrate Australian defence acquisitions," *The Australian Defence Force Journal*, vol. 102, 2017.
- [4] GAO, "Weapon systems cybersecurity: DoD just beginning to grapple with scale of vulnerabilities," United States Government Accountability Office, Report, 2018.

- [5] C. Brown, P. Christensen, J. McNeil, and L. Messerschmidt, "Using the developmental evaluation framework to right size cyber TE test data and infrastructure requirements," *International Test and Evaluation Journal*, vol. 36, pp. 26–34, 2015.
- [6] P. Christensen, "Cybersecurity test and evaluation: A look back, some lessons learned, and a look forward!" *ITEA Journal*, vol. 38, no. 3, p. 221–228, 2017.
- [7] K. Joiner and M. Tutty, "A tale of two allied defence departments: new assurance initiatives for managing increasing system complexity, interconnectedness and vulnerability," *Australian Journal of Multi-Disciplinary Engineering*, vol. 14, no. 1, pp. 4–25, 2018.
- [8] S. Chan, "Prototype orchestration framework as a high exposure dimension cyber defense accelerant amidst ever-increasing cycles of adaptation by attackers: A modified deep belief network accelerated by a stacked generative adversarial network for enhanced event correlation," in *CYBER 2018: The Third International Conference on Cyber-Technologies and Cyber-Systems*. IARIA, Conference Proceedings.
- [9] G. Austin, "Civil defence gaps under cyber blitzkrieg," in *International Conference, Research and Education for the Cyber Storm*. Online, Conference Proceedings, p. 33. [Online]. Available: <https://www.unsw.adfa.edu.au/unsw-canberra-cyber/sites/accs/files/uploads/Cyber%20blitzkrieg%207%20Feb%202019%20CONF%20VERSION.pdf>
- [10] C. H. Heintz, *The Potential Military Impact of Emerging Technologies in the Asia-Pacific Region: A focus on cyber capabilities*, ser. Emerging Critical Technologies and Security in the Asia-Pacific. Hampshire, U.K.: Palgrave Macmillan, 2016.
- [11] L. Purton and K. Kourousis, "Military airworthiness management frameworks: A critical review," *Procedia Engineering*, vol. 80, p. 545–564, 2014. [Online]. Available: <https://doi.org/10.1016/j.proeng.2014.09.111>
- [12] J. P. T. Mo and K. Downey, "System design for transitional aircraft support," *International Journal of Engineering Business Management*, vol. 6, pp. 45–56, 2014. [Online]. Available: <https://doi.org/10.5772/57451>
- [13] R. J. Hodge, S. Craig, J. M. Bradley, and C. B. Keating, "Systems engineering and complex systems governance – lessons for better integration," in *INCOSE International Symposium*, vol. 29, Conference Proceedings, p. 421–433. [Online]. Available: <https://doi.org/10.1002/j.2334-5837.2019.00612.x>
- [14] S. Simmonds and S. C. Cook, "Use of the goal structuring notation to argue technical integrity," in *INCOSE International Symposium*, vol. 27, Conference Proceedings, p. 826–841. [Online]. Available: <https://doi.org/10.1002/j.2334-5837.2017.00396.x>
- [15] K. F. Joiner, "How australia can catch up to U.S. cyber resilience by understanding that cyber survivability test and evaluation drives defense investment," *Information Security Journal: a Global Perspective*, vol. 26, pp. 74–84, 2017.
- [16] M. Thompson, "Towards mature ADF information warfare - four years of growth," Australian Department of Defence, Report, 2020. [Online]. Available: <https://www.defence.gov.au/jcg/Docs/Towards-%20Mature-ADF-%20Information-Warfare.pdf>
- [17] 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)," 29 May 2018 2018. [Online]. Available: <https://parlview.aph.gov.au/mediaPlayer.php?videoID=399539>
- [18] Australian Government, "ADF cyberworthiness governance framework," 2020.
- [19] U.S. DoD, "Cybersecurity maturity model certification (cmmc)," 2021. [Online]. Available: <https://www.acq.osd.mil/cmmc/>
- [20] C. Easttom and W. Butler, "Applying systems engineering principles to penetration testing," in *15th International Conference on Cyber Warfare and Security*, B. K. Payne and H. Wu, Eds. Academic Conferences and Publishing International Ltd., Conference Proceedings, pp. 127–134.
- [21] C. Easttom, "Mathematically modelling cyber-attacks utilizing engineering techniques," in *15th International Conference on Cyber Warfare and Security*, B. K. Payne and H. Wu, Eds. Academic Conferences and Publishing International Ltd, Conference Proceedings, pp. 119–126.