Journal of Information Security and Applications 64 (2022) 103050

Contents lists available at ScienceDirect

Journal of Information Security and Applications

journal homepage: www.elsevier.com/locate/jisa

Assessing cyber threats for storyless systems

Per Håkon Meland[a,c,*], Dag Atle Nesheim[b], Karin Bernsmed[a], Guttorm Sindre[c]

[a] SINTEF Digital, Strindvegen 4, 7465 Trondheim, Norway
[b] SINTEF Ocean, Postboks 4762 Torgard, 7465 Trondheim, Norway
[c] Norwegian University of Science and Technology, Høgskoleringen 1, 7491 Trondheim, Norway

ARTICLE INFO

Keywords:
Cyber threats
Decision-making
Estimation
Empirical evaluation
Case study
Maritime communication

ABSTRACT

A proper assessment of potential cyber threats is vital for security decision-making. This becomes an even more challenging task when dealing with new system designs and industry sectors where there is little or no historical data about past security incidents. We have developed a threat likelihood estimation approach that supports risk management under such circumstances. Quantifiable conditions are determined from the environment in which the system will reside and operate, that is the availability of potential threat actors, their opportunities of performing attacks, the required means that are needed for the attack to succeed, and motivation factors. Our research method follows the principles of practice research where both researchers and practitioners have played central roles in a real-life development project for a maritime communication system. We used a qualitative case study for feature-based evaluation of the approach and associated tool template, and to gather evidence on practical aspects such as suitability for purpose, efficiency and drawbacks from five user groups. The results show that representative participants from the cyber security and maritime community gave positive and consistent scores on the features, and regarded time usage, traceability of the threat assessment and the ability to indicate underlying uncertainty to be very appropriate. The approach has been proven useful for this domain and should be applicable to others as well, but the template requires up-front investments in gathering knowledge that is relevant and reusable in additional context situations.

1. Introduction

Many recent reports show that cyber attacks are becoming more sophisticated and frequent [1–4]. This makes it a difficult task to decide how much and what kind of security is needed to protect organisations and their systems. Cyber security decision-making is uncertain by nature, and even more so when dealing with new system designs and industry sectors that are undergoing rapid digitalisation, opening themselves up to more exposure. Under such circumstances, we can talk about systems that are *storyless*, meaning that there is little or no (his-)story or knowledge related to past security incidents. Such data, e.g., *attack frequency*, *attack type distribution*, *number of successful/prevented attacks*, are often required input when trying to quantify threat likelihood in traditional methods. With storyless systems, we must seek other ways to assess potential threats and their consequences in order to make informed decisions on risk treatment.

The purpose of this paper is to present a systematic approach for assessing threats for storyless systems. The goal has been to develop something that can be readily applied in real-life projects, being efficient in terms of resource usage and flexible enough to be adjusted to the best data available. With this approach, we are able to make threat estimations based on the availability of potential threat actors, their opportunities of performing attacks, the required means (resources) that are needed for the attack to succeed, and motivation factors. Such estimations are less dependent on historical events data, and therefore allow us to use a proactive approach for assessing new designs and prototypes.

Through a case study performed in relation to a maritime system development project, we have sought answers to the following research questions:

1. How can we estimate threat likelihood for a new design?
2. What are the perceived advantages and disadvantages of such an approach?

The project has involved security experts and domain specialists who have participated in actual threat assessments and evaluated the approach. We hope that this contribution will be a practical and relevant addition to existing risk management methods, within the maritime as well as other domains.

[*] Corresponding author at: SINTEF Digital, Strindvegen 4, 7465 Trondheim, Norway.
E-mail addresses: per.h.meland@sintef.no (P.H. Meland), dag.atle.nesheim@sintef.no (D.A. Nesheim), karin.bernsmed@sintef.no (K. Bernsmed), guttorm.sindre@ntnu.no (G. Sindre).

https://doi.org/10.1016/j.jisa.2021.103050

The structure of this paper is as follows. Section 2 provides information about threat modelling, associated concepts, challenges and state of the art. Section 3 explains our research method and case study. Section 4 explains the approach itself with an illustrative example. Our evaluation results are presented in Section 5, and we discuss our results and threats to validity in Section 6. Finally, Section 7 concludes the paper.

2. Background and state of the art

As defined by the ISO/IEC 27000 vocabulary [5]; a *threat* is the potential cause of an unwanted incident, which can result in harm to a system or organisation. When assessing threats, we often talk about *threat modelling*. In 2000, Schneier [6] described threat modelling as a way of imagining the vast vulnerability landscape of a system and ways to attack it. He also made a point that this is something hard to do and only comes with experience. Two decades later, a diverse set of security experts published the *Threat Modeling Manifesto* [7] based on the most common concepts from the literature throughout the years. The manifesto defines threat modelling as “analyzing representations of a system to highlight concerns about security and privacy characteristics”, where some of the most central questions one should try to answer are “what are you building?”, “what can go wrong?”, “what to do about it?” and “did you do a decent analysis job?”.

There is no single, ideal and uniform method of assessing threats and associated risks. There are overarching processes and practices found within standards such as the ISO/IEC 31000- and 27000-series [8, 9] and NIST publications [10,11], but exactly how to perform this will usually depend on factors such as the wanted perspective, experience, personal preferences, available information, and local conditions. When there is little quantitative data available, subjective opinions become central in the assessments. Though security experts and domain specialists can make good estimates on consequences following a cyber event, determining the likelihood factor is a harder challenge as that involves a fair share of guesswork. Böhme et al. have pointed out that [12] “models of cyber risk arrival need to be more predictive.” This is in accordance with Ahrend and Jirotka [13], who state that “cyber security defenders need to make more informed decisions regarding what threats to mitigate and how to mitigate them” and “to do so requires defenders to *anticipate* threat actors’ behaviour”. Almukaynizi et al. [14] have shown a growing community attention towards predicting cyber security events, and argue that predictions should be transparent and interpretable to allow human-in-the-loop-driven decisions.

In the literature we can find different approaches on how to support human-driven predictions of risk factors. For instance, Hubbard [15] has proposed the HTMA approach (*how to measure anything*) for cyber security risks, which heavily relies on subjective expert opinions. Santini et al. [16] have extended this approach, adding more objective data from several sources to progressively improve the risk model. These *key risk indicators* (KRIs) were mainly based on measurements internal to the organisation, such as malware infections, vulnerabilities, data breaches and deep web exposure. Figueira et al. [17] have proposed a mixed qualitative–quantitative risk analysis approach, using regression models instead of data about the past to compute future threat probability. Similar to Santini et al. they base their estimations on currently known system vulnerabilities. Kissoon [18] also applies regression models to measure the effectiveness of current implemented cyber security measures in organisations. She uses internal variables such as risk appetite, security budget and loss after security breach obtained from surveys and interviews. Al-Hadhrami et al. [19] have proposed to use subjective logic based on the criteria vulnerability level and technical attack difficulty to compensate for the lack of accurate, probabilistic data.

The challenge of threat prediction becomes even more apparent with storyless systems, for which there is virtually no data about existing vulnerabilities, attack frequencies or loss after incidents. Our approach is mainly concerned with assessing such systems, and also limiting what is known as *Knightian uncertainty*, where risky (quantifiable) decisions are made based on non-quantifiable conditions [20]. Instead of taking the system-centric view, we determine quantifiable conditions from the environment in which the system resides and operates. Previous work that has been using these premises is for instance presented by Buldas et al. [21], who derive cost of attacks from threat models in order to decide whether the system is a realistic target for gain-oriented attackers. A similar path can also be seen in a series of papers by Knez et al. [22], Llansó et al. [23], McNeil et al. [24], that describe a *capability-based approach* to cyber risk management for space missions. They criticise the amount of labour that is needed to describe attack paths and give likelihood estimation, emphasising that these are too subjective and do not scale well for complex systems. They suggest that mitigations should be based on representations of presumed offencive capabilities of attackers and the defencive capabilities. Recently, ter Beek et al. [25] have developed a framework for quantitative security risk modelling where the cost of an attack (both successful and failed) are calculated and used as a constraint. Similarly, Bagnato et al. [26] use different types of data not tied to past events as part of threat model assessments. They also advocate for the involvement of domain specialists in order to give accurate estimates, and based on a case study they identified so-called conflicting modelling goals that have practical implications on the quality of the risk analysis. These were *time usage* for creating models, *reusability* of context dependent data values, *accuracy* and *simplicity*. Most of these conflicting goals are in line with the later findings from a survey on graphical security models by Hong et al. [27], pointing to common practical challenges related to *scalability* of complex models, *reusability* and *tool availability*.

In most cases we want to make our estimations based on the best data available, which can be a combination of some historical data and subjective opinions. For instance, through a set of case studies, Paté-Cornell et al. [28] have presented several ways to gather and use the information available to quantify cyber risk. For extreme events without data, they suggest using probabilistic analysis of potential scenarios where the limits of statistical data are completed by expert opinions. Examples of data are potential points of access, vulnerabilities, software update time and the costs/loss after successful attacks. Buldas et al. [29] have presented a quantitative attribute approach that deals with incomplete information. This could be applied when there is some historical data and some domain knowledge available to the model.

Related to the maritime domain, Mraković and Vojinović [30] show that regulatory bodies and international organisations set risk assessment as a necessary first step for preventing unwanted events at sea, with several sets of guidelines that refer to the NIST publications. Still, these guidelines do not give details on exactly *how* these assessment should be conducted. Looking at the literature, Tam and Jones [31] have proposed an approach called *Maritime Cyber Risk Assessment* (MaCRA). The risk assessment in MaCRA is based on three dimensions: *system vulnerabilities*, *ease of exploit*, and the *reward* achieved by the attacker. This approach has some similarity to ours: the vulnerability dimension resembles our opportunity factor, the ease-of-exploit dimension resembles our *means* factor (does the attacker have the required means to perform the attack, or at what cost can such means be obtained?), and the reward dimension resembles our *motivation* factor. On the other hand, while our approach has a separate factor for *threat actors*, actors are discussed inside the dimensions of reward and ease-of-exploit in MacRA, for instance, different types of actors (criminals, terrorists, hacktivists) may be pursuing different types of rewards (money, harm to an enemy, attention to political causes), and the ease-of-exploit will be different depending on the type of attacker (e.g. experienced hacker vs. novice). However, the bigger differences are in the way of working with the two approaches. MacRA is based on a pre-cataloguing of different types of actors and target system components typically found in the maritime sector, where picking the system configuration will produce rough estimates of risks for various threats based on historical data. Our approach rather focuses on people working together to produce estimates for the weight of various factors, looking at threats one by one, to arrive at a numerical estimate for the threat likelihood. Hence, rather than being pure competitors, it is also possible that the two approaches could complement each other, using our approach for the estimation of threat values — but with benefits from MaCRA's pre-cataloguing of various system components where applicable, and using a MacRA-inspired approach to visualise the gravity various threats compared to each other in a nice graphical display.

Another work especially addressing maritime cyber-security is Kessler et al. [32], providing a taxonomy to aid risk assessment. The taxonomy supports a way of identifying possible threats to the target system (including both malicious attacks and natural hazards), categorising these threats according to four attributes: the type of attack (e.g., GPS jamming), which security goal (of Confidentiality, Integrity, Availability, Possession, Authenticity, Utility) that this attack would invalidate (e.g., Availability in the case of GPS jamming), which systems are involved (e.g., GPS), and the threat category (e.g., Jamming). Then, estimates of risk for each threat are derived from tables indicating the source of the threat (human attacker or natural hazard), and the likelihood, severity and ease. However, unlike our approach, Kessler et al. do not propose a more detailed support or work process for estimating the values for the likelihood and ease. This is a main difference from our approach, which tries to go in more detail to provide values based on e.g. the attackers opportunities to acquire the necessary means for the attack. Also, our approach does not look at natural hazards, but instead has a more detailed breakdown of various human attackers, assigning weights for various types of attackers.

Svilicic et al. [33] have described how to conduct a cyber risk assessment for a specific ship. The basis of their analysis was a combination of a ship crew survey and a technical vulnerability analysis of some of the ship's critical system components. In contrast to our approach, such an assessment should be more suitable after deployment and when the crew have gained operational experience. You et al. [34] have conducted a literature review on risk assessment methods from other domains. They conclude that these can be easily adapted to maritime and port security, but it is also clear that they will depend on good subjective estimations or historical data.

Further background techniques that our approach directly applies are presented alongside the approach itself in Section 4.

3. Method and materials

Our research method follows the principles of *practice research* as defined by Goldkuhl [35], where both researchers and practitioners play central roles in situational enquiry and generalising knowledge. We have introduced new artefacts in the form of an approach for assessing threats and a tool template that supports this activity. Based on Kitchenham [36,37], we have employed the DESMET evaluation method to assess the appropriateness of our artefacts in the context of a "real" project for the maritime industry. This can be described as qualitative case study, where the evaluators make subjective assessments of the relative importance of different features and how well a feature is implemented. According to Kitchenham, such an evaluation method is suitable when the benefits are observable on a single project and difficult to quantify, and the user population is limited. Zelowitz and Wallace [38] argue that feature analysis is well-suited for evaluating new technology and provide insight into its use, and Marshall [39] has shown that this is an established evaluation method in software engineering. For these reasons we consider feature-based evaluation to be appropriate for our study as well.

As depicted in Fig. 1, we initially developed the approach by combining and adapting existing techniques for threat assessments. Our motivation for doing this was to perform internal risk assessment of the storyless system we were developing as part of our case study project, which required us to document and justify our security trade-offs. As a second step we chose two representative sub-systems to validate the approach, involving security experts and domain specialists that were informally debriefed afterwards. The results of this validation have partly been published by Haga et al. [40]. Though we were able to validate that the needs and expectations were met from the sample of stakeholders, we also saw possibilities for improving the efficiency by reusing some of the model elements and associated values. We therefore expanded the approach and created tool templates to support the activities as part of step three. We now reapplied the approach to a larger set of sub-systems in step four, involving additional stakeholders and performed a more systematic evaluation in step five.

Each evaluation session was conducted as semi-structured interviews, which Robson and McCartan [41] consider most appropriate for researchers who are closely involved with the overall project. We had selected a set of core features that the participants in each session would score according to a Likert scale and comment on as a group. Furthermore, we asked questions recommended by DESMET related to:

- Suitability for purpose — will the overall approach do the job we want it to?
- Is the approach efficient in terms of resource usage?
- Drawbacks — is there any aspect that makes the approach less attractive though it does the job?
- Other advantages — are there other attractive aspects of the approach, beside efficiency and fit for purpose?

All participation was voluntary, and the recorded results were anonymised. The details of the actual threat assessment are confidential, but in the following section we give an overview of the case study system to show the context.

3.1. Case study: A new maritime communication system

The *maritime* domain is defined as "all areas and things of, on, under, relating to, adjacent to, or bordering on a sea, ocean, or other navigable waterway, including all maritime-related activities, infrastructure, people, cargo, and vessels and other conveyances" [42]. According to Kontovas and Psaraftis [43], the *International Maritime Organisation* (IMO) has recognised that the whole philosophy of using historical data for *Formal Safety Assessment* (FSA) cannot be used for new system designs. Furthermore, it is undesirable to wait for new incidents to happen in order to measure the effects of newly implemented risk controls. We believe the same arguments hold for cyber security risks. Though the maritime domain has a long tradition of safety-focus, ENISA [44] has pointed out that the awareness of cyber security in the maritime community has unfortunately been low. At the same time, the domain is characterised by a complex ICT infrastructure with fast technology development.

Although several studies, such as the ones by Caprolu et al. [45], Mraković and Vojinović [30] and Chang et al. [46], give interesting overviews of typical security threats in maritime systems, with some examples of incidents and suggestions of countermeasures, there is little data available to directly quantify the factors relevant for estimating risks. Jacq et al. [47] have proposed a software architecture for monitoring security incidents in maritime systems and setting up a maritime security operations centre to aid vessels in case of attacks. The proposed system would collect data about actual security incidents. If the use of such systems becomes widespread in the future, this would give better data on which to base estimations. Yet at present maritime systems are largely storyless when it comes to cyber-security risk analysis. This yields a need for better support when assessing threats and affirms the domain as interesting from a research perspective.

Our case study has taken place within the context of a research and development project named *Cyber Security in Merchant Shipping Service Evolution* (CySiMS-SE) [48], which lasted from 2019 to 2021. The goal

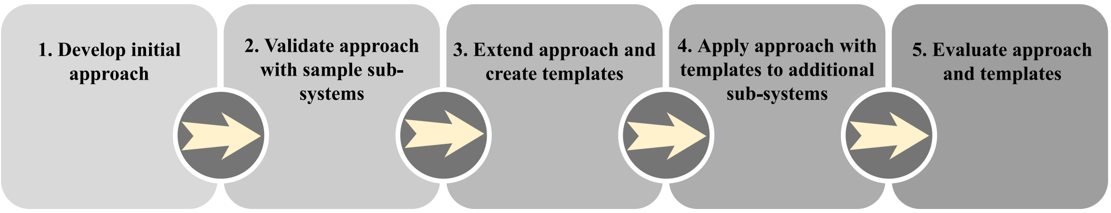

1. Develop initial approach
2. Validate approach with sample sub-systems
3. Extend approach and create templates
4. Apply approach with templates to additional sub-systems
5. Evaluate approach and templates

Fig. 1. Steps for developing and evaluating the approach.

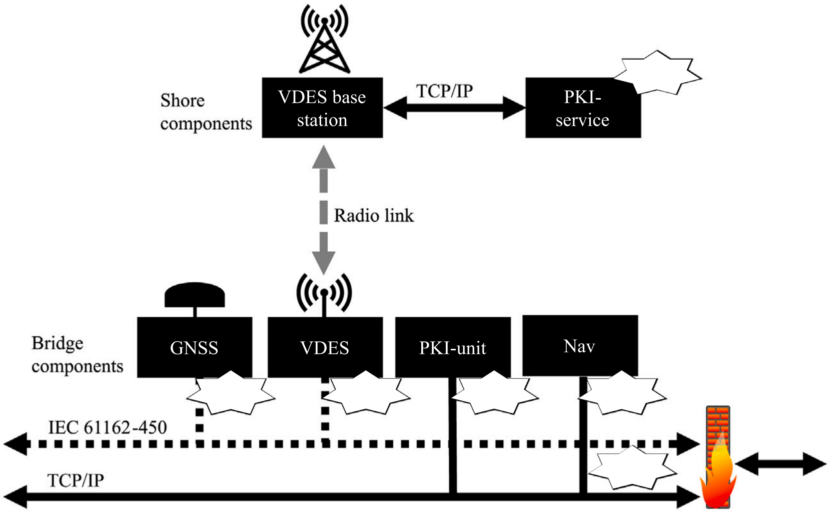


Fig. 2. Threats targeting shore-based and on-board bridge components.

of this project has been to demonstrate and operationalise security for the *VHF Data Exchange System* (VDES) [49] radio and integrate it with the on-board computer architecture. An example use case for this system is for ships to digitally sign and transmit route data to a national coastal administration. A simplified overview of the system is depicted in Fig. 2, which shows the main sub-components and how they are connected. On the bridge of the ship, there is a *Global Navigation Satellite System* (GNSS) providing positioning and time data. The VDES is responsible for data transfer to on-shore base stations. A dedicated *Public Key Infrastructure* (PKI) unit is invoked to perform cryptography functions and securely storing the ship's private key and a cache of public key certificates. A Nav unit integrates digital navigational data and is used by the navigator for planning routes. The GNSS and VDES sub-components are connected to a dedicated IEC 61162-450 [50] compliant network, and traffic needs to go through a firewall to reach either the regular TCP/IP network connected to the PKI-unit and Nav on the bridge or other off-bridge systems, e.g., administrative, crew or entertainment systems. On-shore we can also find a PKI-service that enables enrolment and revocation of certificates, as well as a repository of public key certificates for the flag state.

Based on an analysis [51] of the maritime cyber threat landscape showing that malware infection is the prevalent way of compromising systems, the scope of the assessment has been on the unwanted event that one or several of the sub-components could become infected and the likelihoods associated to this. The threats we have assessed are marked T_{1-5} in Fig. 2, whilst T_0 is used as an example in this paper.

4. The threat likelihood approach explained

This section explains our approach, which should be seen as a customised version of *OWASP Risk Rating Methodology* (OWASPRR) by Williams [52]. Basically, the goal is to "estimate the likelihood of a successful attack from a group of possible attackers" based on a model that is simple to use, yet with enough detail to make accurate estimates. Williams recommends that the risk rating model should be tailored according to specific organisations, and for our approach we have chosen a set of likelihood factors that are more suitable for our use on storyless systems than this reference model.

Fig. 3 shows the four likelihood factors we consider for each threat; threat actors, opportunity, means, and motivation. Since we are dealing with intentional attacks, there will always be threat actors actively involved. The remaining factors are based on the traditional concept from criminal law, that people who commit crime are likely the ones who have *motive*, *means*, and *opportunity* (MMO) to do so [53]. According to Van Ruitenbeek et al. [54], these factors are also applicable for analysis in the cyber realm.

For each factor we apply the threat template to find a weighted value that gives the following indication:

- For *threat actors* the weight indicates how large a group the actor represents in comparison to the other actors.
- For *opportunity* the weight should be based on the threat actor's spatial, temporal and vulnerability exploiting opportunities.
- For *means* the assessment should consider to what extent the different threat actors have the required means needed to perform the attack.
- The *motivation* weight should be based on what motivation factors and intents that can be associated to each threat actor.

The weight values are numerical values between 0 and 10 and we derive the overall threat likelihood value from the average of these.

The threat template provides domain knowledge that supports the estimation of the individual threat factors. The following sections show how to apply the threat template to the example threat T_0 from the maritime case study. The results from each template are used as input

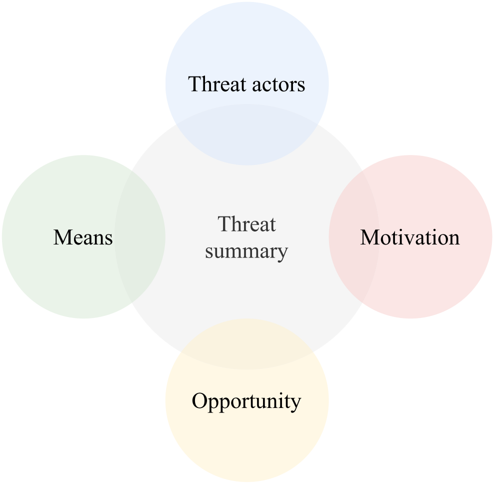


Fig. 3. Threat factors used to derive the overall threat assessment.

to a threat summary, providing traceability and justification for the overall threat likelihood. Just as the OWASPRR, we offer a spreadsheet containing the template and a threat summary. This tool provides documentation of the threat assessment and enables calculation of the numerical values.

Though there is no explicit starting order when working with the different factors, our experience indicates that it is natural to begin with threat actors followed by opportunity, means and motivation. All factors can be revisited and adjusted iteratively throughout the process.

4.1. Identifying threat actors

We use *inductive profiling* [55] as a tool to identify potential offenders before any crime is actually committed. Shinder and Tittel [56] define a profile to be a set of characteristics likely to be shared by criminals who commit a certain type of crime. Our template for threat actors is not only limited to traditional criminals, but also includes relevant actors from the maritime operations who could become involved in a cyber attack.

Fig. 4 shows an excerpt of the taxonomy found within this template. It is not meant to be exhaustive, but serves as an inspiration where the assessors can select, add or join elements that are entered into the threat summary. The actual threat template contains a more thorough description of each actor based on available literature [40,57–61].

Based on the context, we start by picking threat actors that could somehow be involved. In our example we are considering a system component on-board the ship, therefore we include profiles among the crew and can disregard a lot of the actors tied to land-based operations. The relevant actors are marked with a warning sign in Fig. 4.

As with the OWASPRR, we use the weight *size* to indicate how large these groups of threat actors are. The weights between 0 and 10 are not the actual number of people, but values relative to each other. So for instance, with a vessel that has a captain, chief, second and electro-technical officer, these actors are typically given a weight of 1. Alternatively, we could merge them into a more generic officer actor with a weight of 2 – 3. There is usually a slightly higher number of sailors/ratings on-board, which could yield a weight of 4. It also makes sense to apply a weight of 3 for technical workers from the shipping company, who could remotely access components or do physical maintenance on these. Cyber extortionist is given the highest weight, 8, based on the number of potential online cyber criminals we know are out there. Maritime operations are unfortunately often targeted when there are geopolitical conflicts or tension between states. In this example, we assign the weight 5 to cyber warrior as the vessel is sailing under a flag that has a few hostile nations.

4.2. Finding opportunities

Opportunity can be defined as the presence of a favourable combination of circumstances that makes an action possible [62]. Opportunity can therefore be used as an indicator for *when* and *where*,

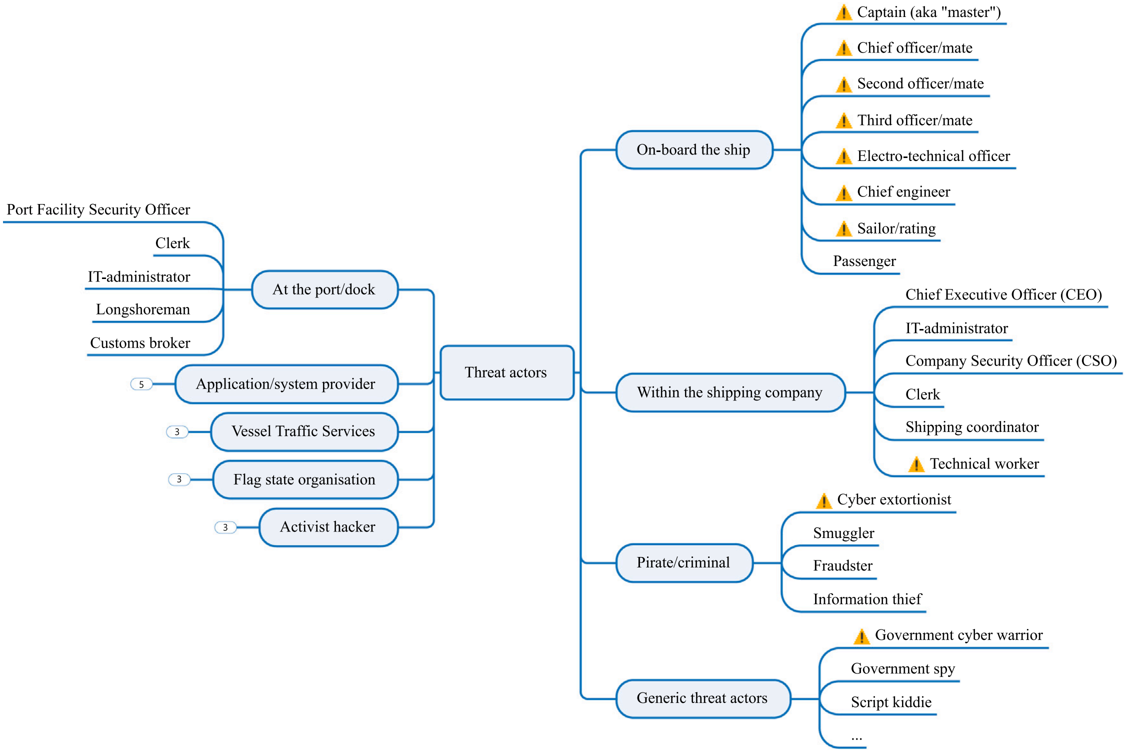


Fig. 4. Potential threat actors found in the template.

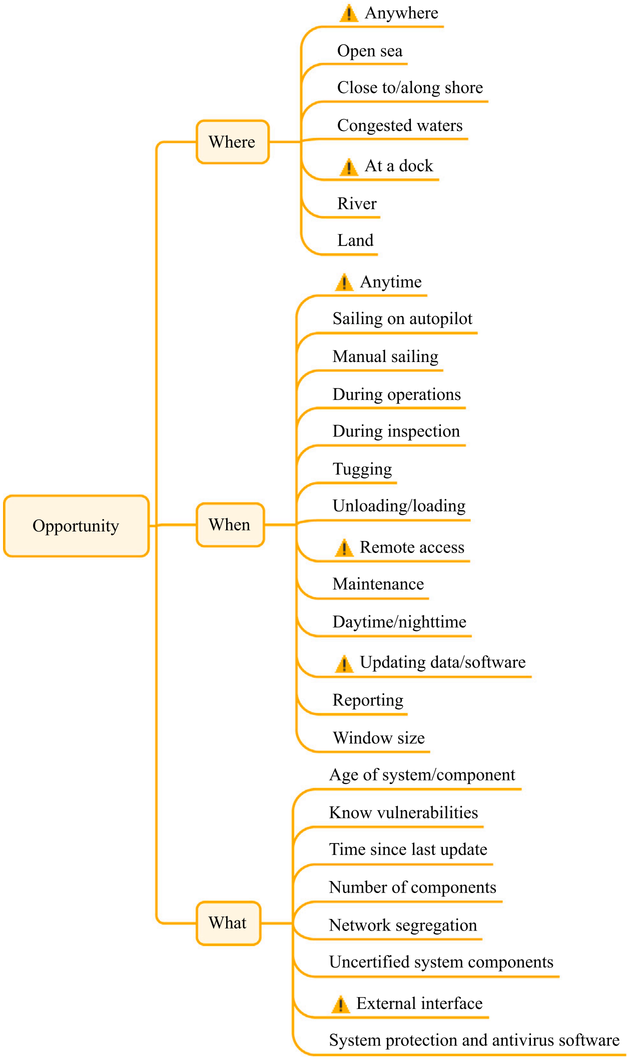


Fig. 5. The template provides potential opportunities for the selected threat actors, divided into circumstances related to where, when and what.

and to some extent *how*, the threat can manifest itself. If there are vulnerabilities that can be exploited from anywhere, at any time, the opportunity weight will be high. If, instead, the adversary must be at the right place at the right time, the weight will be low. In practice, not all vulnerabilities can be eliminated, as this would cause excessive security costs and inhibit meaningful operations. However, we should strive to make the window of opportunity as small as possible so that the adversary cannot easily attack the system without being noticed.

In our threat template for opportunity, we take into account that maritime vessels have a changing operational environment. We have further divided opportunity into three dimensions. The first one is the spatial dimension, which is another name for location. The next opportunity dimension is related to time. In many cases, the spatial and temporal characteristics will be interlinked, for instance sailing on autopilot is usually performed at open sea, while tugging usually takes place in congested waters. It is possible to have several temporal characteristics for opportunity. For instance, a certain attack opportunity may arise while the ship is sailing on autopilot but would need at least 10 min (window size) to succeed.

Our third opportunity dimension is related to system vulnerabilities. There must be such vulnerabilities present in order to exploit the system. Note that many of these indicators are mostly related to legacy systems, and to a lesser degree, new systems still under design/implementation.

Fig. 5 shows an excerpt from the taxonomy found within the template. Based on the context we choose relevant opportunities (marked with a warning sign) for the threat actors and provide a weight with a justification in the threat summary.

4.3. Deriving the necessary means

The required means or resources needed to perform an attack is another factor that helps us determine the threat likelihood. While cheap attacks can potentially be implemented by many, more expensive ones require attackers that are more determined to invest. As shown in Section 2, there are different approaches for estimating attacker costs, however, most of these are based on known attack paths. With new designs it is more difficult to predict attack paths.

We utilise an approach described by Haga et al. [40], which again is based on two methods with an already high uptake in the security community, namely the Cyber Kill Chain by Lockheed Martin [63,64] and attack trees by Schneier [65]. Here, a resource tree can be modelled for each consecutive stage of a cyber-attack. These trees estimate the fundamental resources that are required to complete this stage and move on to the next one, but differ from traditional attack trees since they are not concerned about the details of the attack paths. The tree consists of a root node, defining the cyber kill stage, a second level of conjunctive resource classes, and a third level of disjunctive resource alternatives. We assign monetary cost values for the resource alternatives along with an optional confidence value. For instance, if the attacker would require a certain type of hardware to perform the attack, and the direct cost of that item is known, we can assign that value with a confidence value close to 1 (certain). However, in cases where we are unsure about the cost, for instance for finding exploitable vulnerabilities, we use a low value such as 0.2 (uncertain). The cost and confidence values propagate up the trees from the included kill chain stages.

Our means template is an alternative to the *Interactive Resource Cost Model* (IRCM) tool by Haga et al. [40]. Instead of having to model the resource trees from scratch, generic structures are part of the template and only need cost values and optionally confidence. These structures were developed from the validation phase, as we saw that there were a lot of common tree elements in the models created for the sample sub-systems. While Haga et al. [40] operate with cost intervals for the resource alternatives, our means template simplifies the estimation task by propagating the minimum expected costs (α) from the alternatives (V) for each required resource (R_j). The total estimated minimum means (M) is the sum of all required resources from the included kill stages, which can be formally expressed as:

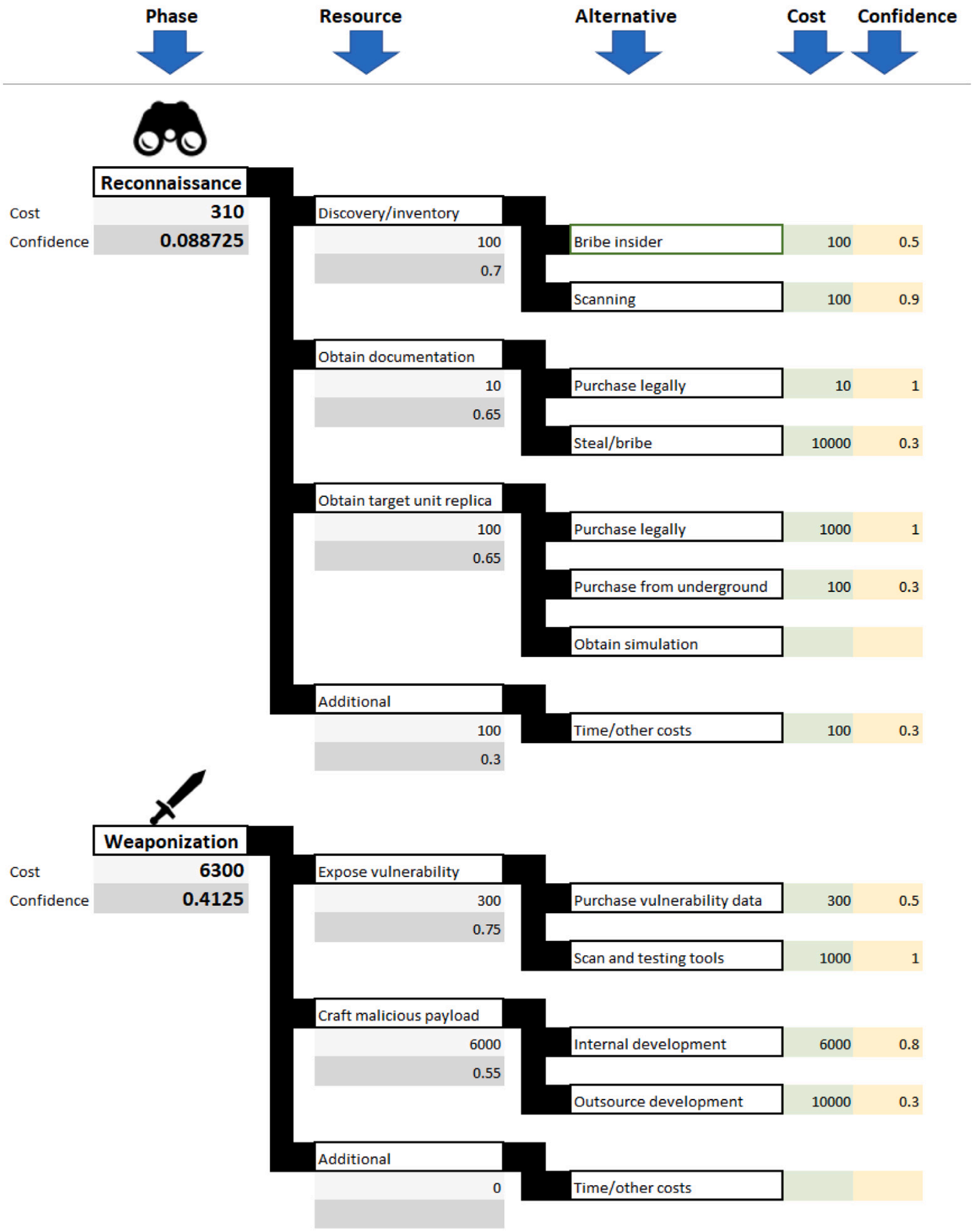
$$M = \sum_{\substack{\text{stage } \in \\ \text{kill chain}}} \sum_{i \in V} \alpha_i \tag{1}$$

As suggested by Haga et al. [40], the overall confidence (C) is the product of the average confidence of the resource alternatives (c_i) to all resources (R_j) for the included kill chain stages:

$$C = \prod_{\substack{\text{stage } \in \\ \text{kill chain}}} \prod_{R} \frac{\sum_{i \in R_j} c_i}{n} \tag{2}$$

Fig. 6 shows a screenshot excerpt from the means template applied to T_0, involving the reconnaissance and weaponization kill stages. Where resource alternatives or stages are considered irrelevant for the assessment, the cost cells can be left blank. Blank confidence values are treated as 1 unless specified otherwise.

An essential part of reconnaissance is to do discovery on the target system, meaning to gain knowledge about which components/software are installed. This kind of information could for instance be obtained



Phase	Resource	Alternative	Cost	Confidence
Reconnaissance Cost 310 Confidence 0.088725	Discovery/inventory 100 0.7	Bribe insider	100	0.5
		Scanning	100	0.9
	Obtain documentation 10 0.65	Purchase legally	10	1
		Steal/bribe	10000	0.3
	Obtain target unit replica 100 0.65	Purchase legally	1000	1
		Purchase from underground	100	0.3
		Obtain simulation		
	Additional 100 0.3	Time/other costs	100	0.3
Weaponization Cost 6300 Confidence 0.4125	Expose vulnerability 300 0.75	Purchase vulnerability data	300	0.5
		Scan and testing tools	1000	1
	Craft malicious payload 6000 0.55	Internal development	6000	0.8
		Outsource development	10000	0.3
	Additional 0	Time/other costs		

Fig. 6. Tool screenshot of the means template, which takes attacker cost with confidence values as input to the various kill stages.

from someone on the inside or using more technical scanning techniques (querying external interfaces or analysing data packages). In this example both of these options have a similar cost estimate of $100, but since we are more unsure about how easily an insider would give up the information, the confidence value is set to 0.5. Since both values are the same, the cost of the discovery/inventory resource amounts to $100, while the confidence becomes 0.7 (average).

An attacker would also have an interest in obtaining documentation of the target system, and that could be done legally at a relative low cost for this particular GNSS component. We can actually find and purchase the documentation from the system provider Web-side, which means an accurate cost estimate with a high confidence. The other alternative is to obtain the documentation in an illegal way, for instance by breaking into the system provider premises or bribing an insider. Since it is the minimum cost that propagates up the tree, it does not matter so much which cost we put into this alternative as long as it is higher than the one above. After a discussion with the system providers, who know their premises and employees best, we assume a sum of at least $10000, but with a low confidence.

Another typical part of reconnaissance is to obtain a target unit replica that the attacker could test and experiment with. In some cases, the target component could simply be purchased directly from the supplier for a known cost, in this example $1000. It is often possible to obtain a unit from underground channels, black markets, or online auctions. In the GNSS example we can quickly search sites such as ebay.com to get price listings of similar second-hand units. Since it is more difficult to know the state of used components, possibly stolen from a ship recycling facility, we have set the confidence to 0.3. If a physical unit is not needed, another alternative would be to obtain simulation software. However, since we already know that the underground alternative is so cheap, we do not have to spend time on this estimate. We can also add additional cost to the reconnaissance stage for expenses we cannot fit under the template structure.

The weaponization stage represents the resources an attacker would have to invest in order to find exploitable vulnerabilities in the target

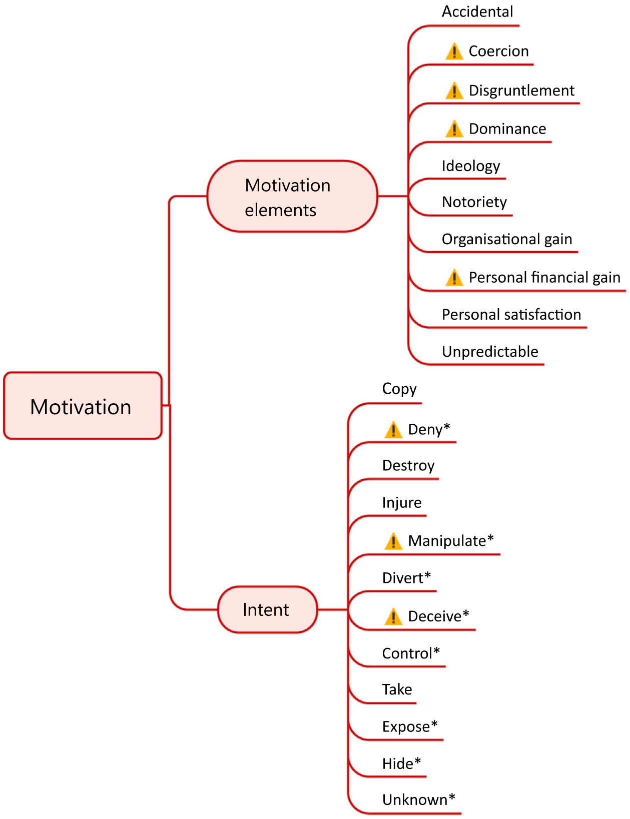


Fig. 7. The template suggests possible motivation factors and intended actions that could be tied to the threat actors.

system and craft a malicious payload. The threat template contains some reference values that can be of support when making these estimates. This includes typical prices for vulnerability data as announced in darknet fora and marketplaces (see e.g. Meland et al. [66]), average size of malware (from Calleja et al. [67]) and average development costs per *source line of code* (SLOC). These numbers are used as a starting point when discussing with system owners what kind of investment would be needed to make malware that could perform an exploit. We also include reference values for outsourcing development based on hacker group ads as a basis for discussion. Of course, the costs of weaponization are crude estimates, only meant to roughly indicate the magnitude of attacker investment.

After the threat template calculates the resulting means value and confidence, we have to create weights for the threat summary. For each threat actor we consider how likely it would be to obtain the required amount of resources. A weight value of 1 indicates that it would be nearly impossible for the threat actor, while the other end of the scale implies that the resource costs are insignificant.

4.4. *What are the motives and intent?*

Motivation identifies the driver that causes the threat agent to commit harmful acts, and we employ the taxonomy by Casey [68] in our motivation template to help us identify the nature of the expected harmful actions. This taxonomy is shown in Fig. 7, and as the motivations are independent of each other, we can assign any number to one or several of the threat actors. A concept related to motive is *intent*, which in criminal law is concerned with the purposeful action the threat actor is willing to carry out [69]. We have extended the objective actions presented by Casey [70] with what we consider to be additional relevant intents (marked with *).

Based on the motivation template we discuss and fill in values for each threat actor in the threat summary with a justification of our selection. Just as with the other likelihood factors, we assign a weight between 0 and 10 by considering what the actor will get out of it if the attack succeeds (*reward*). Similar to motive in the OWASPRR [52], a weight close to 0 indicates that there is little or no reward, a value around 5 possible reward, and 10 a high reward.

4.5. *The overall threat and unwanted event estimation*

Having completed likelihood estimations for threat actors, their opportunity, means and motivation, we are now ready to make a combined average weight as shown in Table 1. In this example there are many possible threat agents, of whom cyber extortionist has the highest average weight (6.25), which we will use as the overall likelihood for this threat. As pointed out by Williams in the OWASPRR [52], it is better to “err on the side of caution” and use the worst-case threat agent and that likelihood value.

Our example threat (T_0) is one of the possible threats that can cause an unwanted event, as seen in Fig. 8. The model in this figure is a bow-tie diagram [71–73], which is one possible way of graphically representing multiple potential threats and consequences. It was applied in our case study since this notation is well-known from risk management within the maritime industry.

In order to give an overall threat estimation that can be utilised in a risk assessment, we can for instance apply the model for combining mutually independent threats as proposed by Bernsmed et al. [74]. It is straight forward to normalise the likelihood values of the threats to probability values by dividing by 10. Given the assumption that the threats can manifest themselves as cyber attacks independently, the probability of the unwanted event U can be computed as:

$$p(U) = p\left(at\ least\ one\ T_i\ occurs\right) = 1 - \prod_{i=1}^{n}\left(1 - p\left(T_i\right)\right) \tag{3}$$

where $p(T_i)$, $i = 1 \dots n$, is the probability of threat T_i.

According to Bernsmed et al. [74], Eq. (3) is much more realistic than simplistic models where threats are considered mutually exclusive (i.e. $p(U)$ will be computed as a sum of the individual threats). Allowing threats to manifest themselves within the same time interval corresponds more closely to the real world, where multiple attackers can work simultaneously to exploit different vulnerabilities.

In our case we end up with a probability for the unwanted event close to 0.96 when we apply Eq. (3) for $T_{0..5}$ with the example likelihood values from Fig. 8. We would subsequently try to assess the risk by taking consequences ($C_{1..3}$) and treatments into consideration as well. However, this kind of continued risk assessment has been outside the scope of this study and evaluation.

5. Evaluation results

Step 4 and 5 of Fig. 1 were conducted in five separate workshop sessions assessing the threats $T_{1..5}$ (see Section 3.1 with five groups of participants, G1-5). The configuration of these groups is shown in Table 2, showing the distribution of security experts and domain specialists among the participants. One security expert acted as an overall session facilitator and one domain specialist was responsible for taking observational notes and record statements during all the sessions, whereas the rest of the participants belonged to the owner (organisation) of the component that the given threat was targeting. The organisations had first-hand knowledge of their own components and operations, with prior experience from assessing risks towards these and similar systems using various techniques. Though the organisations originate from the same geographical area (Norway), they are all well-recognised in international shipping and provide systems and services to customers globally. The results included in this paper do not contain any information that promotes or discredits these. Furthermore, the

Table 1
A simplified threat summary.

Threat actor	Weight	Opportunity	Weight	Means assessment	Weight	Motivation (intent)	Weight	Average weight
Officer (multiple types)	3	Anytime, anywhere	8	Lower required means than the reference value, but still significant.	5	Coercion, personal financial gain, accidental (manipulate, deceive).	3	4.75
Sailor/rating	4	Anytime, anywhere	5	Significant sum for this kind of crew.	3	Coercion, personal financial gain, disgruntlement (manipulate, deceive).	5	4.25
Technical worker	3	At a dock, updating	7	Already has expertise and resources available, lower required means than reference value.	5	Coercion, personal financial gain, accidental (manipulate).	3	4.5
Cyber extortionist	8	Remote access, external interface	4	Experience from similar attacks would lower required means.	5	Personal financial gain (deny).	8	6.25
Government cyber warrior	5	Remote access, external interface	4	Unlimited resources.	3	Dominance (deny, manipulate, deceive).	5	4.25

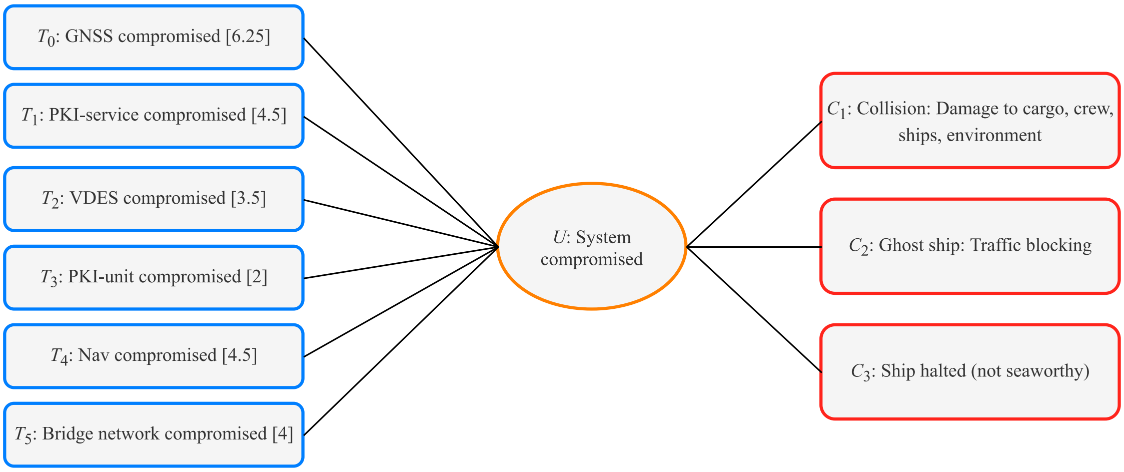

T_0: GNSS compromised [6.25]
T_1: PKI-service compromised [4.5]
T_2: VDES compromised [3.5]
T_3: PKI-unit compromised [2]
T_4: Nav compromised [4.5]
T_5: Bridge network compromised [4]
U: System compromised
C_1: Collision: Damage to cargo, crew, ships, environment
C_2: Ghost ship: Traffic blocking
C_3: Ship halted (not seaworthy)

Fig. 8. A bow-tie model showing different threats that can cause an unwanted event and subsequent consequences. Likelihood values shown in brackets.

participants had no commercial nor conflicting interests related to the threat modelling approach. Four of the participants had experience from the validation of the initial version of the approach, and all had a general awareness of it since it had been developed as part of the CySiMS-SE [48] project that they had participated in.

Each session was organised online using video conferencing, lasted between 60 and 90 min and was conducted in Norwegian, as this was the native language of all participants. To ensure a proper mindset for the participants, there was a general introduction to the session explaining the goals and restrictions of the evaluation. Afterwards, a summary of the results was sent to all participants, so that they could comment, modify and finally approve these contents.

5.1. Feature-based evaluation results

As already explained in Section 3, we applied a feature-based evaluation. The features we selected correspond to the four likelihood factors for threat actors, opportunity, means and motivation, as well as finding the overall threat estimation value based on these. The participants discussed how well the approach and templates supported the determination of these estimation values, and agreed upon a score from a Likert scale between −1 and 5 described in Table 3. The resulting scores from each group for each feature are shown in Fig. 9. In general, we obtained positive scores for all features, with little variance for each group of participants, but more interesting are the comments and suggestions we recorded from the discussions. The following sections give a summary of these comments and our interpretation of their significance.

5.1.1. Identify potential threat actors

This feature received the highest average score (3.8), which indicates a very strong support of the approach. The rather extensive list of potential threat actors found within the template was considered to be a very good starting point for the participants' selections. One of the participants stated that "this is a systematic approach for assessing threat actors. It cannot be trusted 100%, but it's a good basis for further discussion." Other statements were: "you still need to think for yourself, but this support is appreciated", "helps set the mindset for the threat picture" and "the template saves us a lot of time". A suggestion from one of the participants was that "the taxonomy could be linked to what the maritime industry already considers to be the prevalent threat actors".

Based on our observations, we believe that the level of exhaustiveness must be a compromise between completeness and effectiveness for the assessment itself. It requires steady guidance from the facilitator to ensure that time is not wasted on discussing minor or less relevant threat actors. For all groups, several threat actors that were similar in nature were merged into fewer to avoid repetition and save time.

It was also observed that some participants found it difficult to discuss potential threat actors when the context of the assessment was too vague, e.g., that the details of the ship, cargo and operations were not specific enough. This context information could have been used to reduce the taxonomy to begin with, for instance by removing *passenger* for cargo ships.

Furthermore, some participants found it somewhat difficult to discuss potential threat actors without relating these to the foreseen

Table 2
Participants in the evaluation.

Threat	Group	Organisation	Security experts	Domain specialists	Total participants
T_1	G1	Maritime authority	2	3	5
T_2	G2	System provider	1	1	3
T_3	G3	System provider	3	1	4
T_4	G4	System provider	1	2	3
T_5	G5	Maritime research	1	2	3

Table 3
Likert scale definitions adapted from Kitchenham [37].

Generic scale point	Definition of scale point	Scale point mapping
Makes things worse	Cause confusion. The way the feature is implemented makes it difficult to use and/or encouraged incorrect use of the feature.	−1
No support	Fails to recognise it. The feature is not supported.	0
Little support	The feature is supported indirectly, for example by the use of other tool features in non-standard combinations.	1
Some support	The feature appears explicitly in the feature list of the tools. However, some aspects of feature use are not catered for.	2
Strong support	The feature appears explicitly in the feature list of the tools. All aspects of the feature are covered but use of the feature depends on the expertise of the user.	3
Very strong support	The feature appears explicitly in the feature list of the tools. All aspects of the feature are covered and the tool provides tailored dialogue boxes to assist the user.	4
Full support	The feature appears explicitly in the feature list of the tools. All aspects of the feature are covered and the tool provides user scenarios to assist the user such as "Wizards".	5

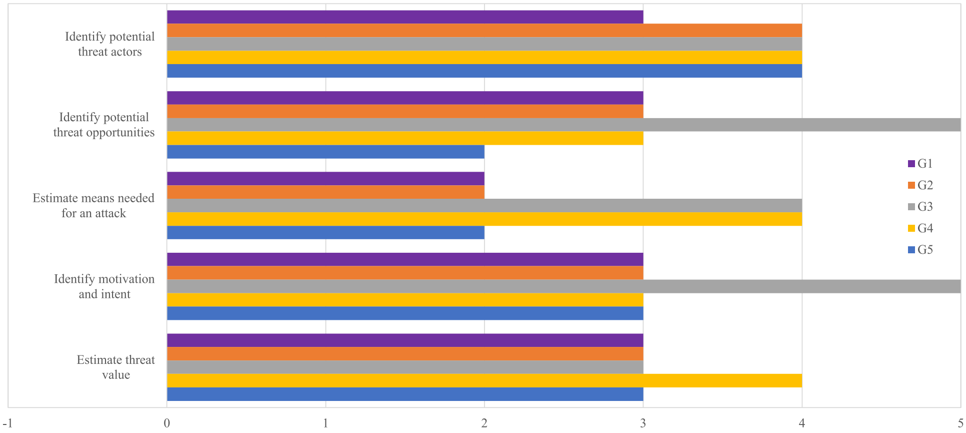


Fig. 9. Scores from the feature-based evaluation.

barriers implemented to mitigate threat actors' access to the asset(s), and the threat actors' motivation and intent to instigate an actual attack. These issues were more of a concern in later stages of the sessions related to opportunity and motivation, which the facilitator explained to the participants. By shifting between or iterating through the different templates we could in practice use, e.g., motivation as a screening criterion for the threat actors as well.

Finally, the concept of weight size spurred some confusion among participants. The facilitator had to point out that we were looking for relative and not precise numbers for the given threat actors. For the template, we may benefit from creating a standardised presentation of the size parameter with concrete examples from the industry (for instance the number of crew on-board certain ship types and/or ship sizes).

5.1.2. Identify potential threat opportunities

This feature had an average score of 3.2 indicating strong support from the approach. The statements from the participants were among similar lines, for instance that "the template has suitable content", "I could not think of anything that was not already there" and "it kick-starts the reasoning process". At the same time, it was expressed that it provides "somewhat lower support (than threat actors), I'm not sure we have caught every aspect".

We noted that all participants expressed a need to identify potential threat opportunities. Nevertheless, the concept of *where* was considered less relevant than *when*, possibly because some participants related cyber threats to remote access attacks only, hence considering physical attack points as less relevant. We do not think this was a major issue for the assessment, but a lesson learned is that the approach would benefit from an improved explanation of the importance and implications of the *where* concept.

The concept of *when* was considered highly relevant in some attack situations (such as disabling ship navigation in ports or high traffic areas), but also less relevant for other types of attack (such as stealing or denying access to information). As with threat actors, there is a need to ensure a proper compromise between being generic and specific for our assessment.

The concept of *what* was not considered very relevant in this case study since we were assessing new designs, though the storyless characteristics of the systems, such as number of components (complexity), network segregation and external interface could have been highlighted more during the opportunity discussions.

5.1.3. Estimate means needed for an attack

This feature received the lowest average score (2.8), and it is also the activity within the approach that requires most time and effort. The feedback from the participants indicated pros and cons for this part of the approach, such as "the template saves us a lot of time coming up with estimates, but it is still a difficult task. The confidence parameter is important", "this is a cool way of calculating attack costs, which is not tied to a specific attack ... at the same time we lost track of what we were really trying to achieve" and "it would have been difficult to estimate attack costs without the template, we do not have a clear idea about these costs to begin with". There were also suggestions for improvements, though the participants acknowledged that this would require more effort, for instance "ideally we should estimate costs for each of the selected threat actors, but that would be too time consuming". Another participant suggested to reduce effort at the cost of accuracy: "we could perhaps simplify the template by using scales rather than explicit costs, the estimations will be rough anyway".

The sheer size of the template puts substantial demands on the facilitator in terms of guiding the participants through the different phases of planning and executing an attack. However, we observed good practices of reducing the scope, such as disregarding the most "mission impossible" inspired ways of attacking. In addition, the option of skipping or de-emphasising some of the attack phases enabled a more practical approach that can be adapted to the most likely attack scenarios. All the threats in our case study were related to malware infections, and the groups focused mainly on the *reconnaissance*, *weaponization* and *delivery* phases of the cyber kill chain [64]. In these phases, there are typical direct costs that the participants could relate to, while in the later phases the main means are more about effort or indirect costs.

It was commented that good estimates require a combination of industry domain and ICT/security knowledge, which was regarded as well-balanced in our groups. However, it was an important task for the facilitator to keep the details of the discussion to a level that everyone could relate to. Also, searching for second-hand maritime technology on ebay.com and other market sites seemed to be a fun exercise to get price estimates, but could also steal quite some time and focus from the assessment. The use of USD as standard currency evoked some unnecessary confusion as this was a foreign currency for the participants. It may be beneficial to use the local currency or automatically convert currencies on the fly. This was no deal-breaker, but in some cases the trail of thought was broken and extra time had to be spent to align the amounts.

Finally, it became somewhat evident from the template that the relationship between blackmailing and bribing is something that must be considered depending on crew and location. Shipping is an international industry where crew originate from all over the world. In low-cost countries, a bribe may be cheaper than blackmailing, while in high-cost countries, the situation may be opposite. One could also relate this to cultural differences, but such a minefield may be better to avoid for the sake of the discussion.

5.1.4. Identify motivation and intent

This feature had a high average score (3.4) between strong and very strong support. From three of the groups there was a general agreement that the taxonomy of motivation and intent seemed adequate, while one group stated that "maybe it is more complete than necessary".

Participants saw this feature as very relevant and as useful documentation in addition to just determining a numerical weight value. Nevertheless, the role of motivation and intent, and especially their interrelationship, were observed to be somewhat confusing at times. One may argue that motive is more closely linked to the threat actors and should be part of their identification. Intent on the other hand, is more an aspect of the attack or its consequence, and was a subject that also came up when discussing means. The facilitator needs to guide these discussions and possibly shift between different parts of the template if new aspects are identified, e.g., an additional threat actor based on discussion around motivation. Also, the sheer number of motivational elements and intents require steady facilitation to ensure that focus is kept on the most relevant ones.

5.1.5. Estimate threat value

The feature that summarised the results from the other estimates received an average score of 3.2, indicating a strong support. It derives weighted values for threat actors, opportunity, means and motivation seen in combination with each other, calculates an average weighted threat value and highlights the most likely threat actor. It was stated that it "provides good background documentation of the estimates and basis for decision-making". Another participant pointed out that it "provides a good structure and ranking of threat actors, but could also lead to a false sense of completion. The approach is good as long as the implementation (of it) is done properly". As each group only assessed one type of threat towards their component, they could not really see the greater threat picture. This became apparent by the statement: "we cannot really say what the threat value means without knowing the other threats".

In general, all participants expressed positive remarks towards how the different stages in the approach resulted in an overview. It is imperative that we have identified which threats to include in the assessment in the first place. Even threats with a low score, e.g., T_3, are still relevant and should by no means be disregarded. When we apply Eq. (3), such threats contribute to raising the overall probability of the following unwanted event. This implies that with more threats, the more likely the unwanted event becomes. At the same time, assessing many threats is time consuming and we would like to include the ones that really makes an impact to the probability of the unwanted event, and subsequently a quantifiable risk value when we also take consequences into account.

5.2. Evaluation of the approach as a whole

The last part of the evaluation treated questions from DESMET related to suitability for purpose, efficiency, drawbacks, and other advantages as mentioned in Section 3. Though these answers partially repeated or overlapped with the feature-based answers, the sections below summarise the participants opinions on the approach as a whole.

5.2.1. Suitability for purpose

Our impression is that the participants regarded the approach as a suitable tool for assessing threats. This was backed by the statements: "(the approach) achieves what it's meant to achieve", "it does what it's supposed to do in a good way" and "this is a scientific approach that both reduces and shows uncertainty. It would have been more difficult to estimate threat likelihood without this kind of organisation".

They also saw it as a useful addition to more classic (and more resource demanding) methods for estimating threat likelihood based on threat intelligence and historical data. Some participants even saw it as better than classic methods as the data availability, or the lack of thereof, is a barrier when trying to use statistical probability. It was stated that "risk assessments are notoriously difficult, and anything that helps is a step in the right direction. This approach utilises several (likelihood) factors, which gives more credibility to the result".

It takes some time to become familiar with the approach and the threat template, even for the people involved in developing these. We believe this will improve with time and application, something that was expressed by one of the participants as well: "it's a good tool, but we need more experience with it".

5.2.2. Efficiency in terms of resource usage

The participants from all groups shared mostly positive responses related to the time invested in the assessments, such as “it is pretty effective ... not sure the results would have been different if we spent more time”, “I don’t think we would get better results if we spent a week on this”, “with other methods it would have been difficult to get just as good answers in shorter time” and “it is much more efficient to use the template than creating models from scratch”.

The participants seemed to think that the approach was relatively simple to use, and yet there is some flexibility on how much time and effort that could be spent for each likelihood factor. Less time usually means less details, so there is always a trade-off. It was stated that “it’s a good thing that we do not model specific attacks. That’s complicated and expensive to do, and this approach provides just as good prioritisation of potential threat events”.

Based on our observations, 60-minute sessions would probably be too short for the type of threats we assessed as part of our case study, while 90 min proved to be more suitable.

5.2.3. Drawbacks

Though the approach seemed to do the job it was designed for, there were also some weaknesses pointed out. For instance, there were statements related to presence of uncertainty, but without clear suggestions for improvements: “even with this approach there is still a good deal of gut feeling, which is hard to quantify. However, the same issue goes for all other methods as well”, “some of the likelihood factors are easier to assess than others. The approach has great potential, though we have to accept that there is still a lot of uncertainty. I’m not aware of other methods that are more practical” and “the baseline information within the template, how complete is that?”.

We also recorded more detailed comments on the contents of the threat assessment, such as “opportunities related to physical access to the system could have been better explained. Maintenance (crew) would often have full access, but that would be logged and misuse detected. The model did not represent this in a clear way”. It should be noted that taking risk modifiers into account were not really the goal of this assessment. At the same time, it may be unnatural to discuss threats without considering existing barriers in the system environment.

One minor remark that should be easy to fix was “the terminology should have been translated (to Norwegian) to avoid some confusion and ease the discussion”.

All in all, it seems like the main drawbacks are not unique to this approach, and it would benefit from being adjusted to the local context.

5.2.4. Other advantages

This discussion point revisited many points that had already been covered, such that the approach “gives a quantification of uncertainty, which is a great plus” and “provides an insight into the underlying details/factors”. A bonus effect that could be highlighted was that the participants thought the approach bridged the communication gap between the domain specialists and ICT security experts. It was stated that “in a way, the discussions are useful by themselves”, and that it is useful to get these groups talking together as early as possible in such a project.

6. Discussion

We have developed the threat likelihood approach and associated template as artefacts addressing our first research question; *how can we estimate threat likelihood for a new design?* It should not be seen as a total replacement for existing assessment practices, but as an additional, systematic aid when dealing with storyless systems, that may still be on the drawing board or have not been released into the wild yet. At such stages, there is little quantifiable data such as known vulnerabilities, expected attack frequencies, and malware infections, which are often required input to traditional threat or risk analysis. Instead, threat likelihood estimates are based subjective predictions from security experts and domain specialists, coupled with quantifiable conditions derived from the system environment.

We have also tried to address some of the challenges related to practical application of such techniques, as shown by Bagnato et al. [26] and Hong et al. [27]. First and foremost, the amount of work put into detailed analysis of all possible attack opportunities can quickly outgrow its usefulness. Therefore, we have sought to develop an approach that is efficient but still accurate enough for its purpose. The level of detail should be adjusted to the need of the estimation task. One might want to drill down thoroughly for certain threats, which requires more effort than giving a superficial estimate for threats that are already well-known. For similar threats, it might be sufficient to do a detailed analysis of one and use those results for the others. The approach is based on a number of existing techniques and concepts, such as capability-based risk management [22–24], resource-cost modelling [40], the OWASP Risk Rating Methodology [52] and means, motive, and opportunity from criminal law [53]. Hence, it should be seen more as an evolutionary than revolutionary approach, with flexibility to be combined with other methods and techniques as well.

Creating a template for a specific domain, in our case maritime communication, is another way of achieving more efficiency and accuracy, but this requires a substantial up-front investment that can only be justified if it can be re-used for a large enough set of assessments. For our case study, this has already proved to be worthwhile as we have been assessing several systems more than one time within the same domain. The template [75] has been made openly available under a CC BY 4.0 license and can be readily applied to similar projects, thus seeking to address the tool availability challenge mentioned by Hong et al. [27].

In order to address our second research question; *what are the perceived advantages and disadvantages of such an approach?*, we have performed qualitative evaluations with domain specialists and security experts from our case study. Section 5 has already provided the main findings from the evaluation of the features and overall approach. The feature of identifying potential threat actors along with their relative size parameter was very well received by the participants, while finding opportunities, motivation and overall threat value were also considered as strongly supported. Estimating the means needed for an attack was the most demanding task in terms of time usage and finding quantifiable values, and received a score somewhat lower than the others. Still, this was a clearly positive score and the statements from the participants indicate that they liked the method despite being unfamiliar with it.

The second part of the evaluation confirmed many of the positive remarks that already had been given for the features, and both suitability and efficiency were highly valued. We did not perform a direct benchmark comparison with any specific alternative methods, which would have required a different evaluation setup. However, the security experts and domain specialists were familiar with various types of assessments methods from before, so the statements related to time usage and drawbacks should be seen as a general comparison. It is noteworthy that beside providing threat likelihood with traceability, the approach also worked as a platform for discussion that the participants appreciated. This shows the importance of having some common ground where people with different expertise can interact.

Based on the evaluation, we believe that the approach and template can become even more appreciated with some slight adjustments and increased familiarity among its users.

6.1. Threats to validity

As argued by Cruzes and ben Othmane [76], there will always be a number of potential threats to validity related to science of security and empirical software engineering. However, there are ways of mitigating these threats and thus improving the quality of the research. Here, we highlight threats related to *credibility*, *transferability*, *dependability* and *confirmability*.

Making use of an established evaluation method increased credibility and ensured that we gathered both supporting and discrepant opinions and observations concerning the approach. As depicted in Fig. 1, we had developed a self-conscious research design that followed our case study project. Such a prolonged research engagement allowed us to do early validation, try out alternative variations within the approach and gave the researchers an opportunity to build trust with the end-users. At the same time, the threat assessment was only one of the tasks performed within the overall project, and most of the attention targeted the specification, implementation, and testing of the communication system itself. This gave the case study a realistic context where the approach was used in practice for security decision-making related to ongoing development. The results gave the participants a direct benefit and was not seen as an irrelevant extra burden. We have tried to address the bias of convenience sampling by making sure that the participants had different backgrounds and belonged to different types of organisations (see Table 2), but we acknowledge that the population was rather small. This limitation was the main reason why we chose a qualitative case study evaluation to begin with.

Though the approach was applied within a maritime cyber threat context, there are reasons to believe that it may be transferable to other domains and projects as well. First, the approach is based on existing techniques and concepts that have to some extent already been applied and evaluated for other domains. These techniques also come with some of their inherent limitations. For instance, the cyber kill chain has been criticised for being too much focused on malware, not capturing other types of attack so well. Pols [77] has shown that to remedy this limitation, the literature suggests many variations of the kill chain, some with up to eighteen different phases. For our approach, there is flexibility on which and how many phases to include, but as already mentioned in Section 5.1.3, it was for the first three phases that the participants could most easily estimate concrete costs. Second, we have provided a narrative context description as part of Section 3 to make it easier for other researchers or practitioners to judge whether the approach would fit for application partly or as whole in other assessments. Third, many of the participants had solid backgrounds from other domains, and were thus able to give opinions on transferability and external validity.

Based on the consistency of the scores from the feature-based evaluation (see Fig. 9), we argue for a certain extent of dependable results from the evaluation. It is more difficult to assess the dependability of the threat assessment itself, since the different groups had their own sub-component as the main scope. Since the actual results of these assessments are confidential, we are unable to show what the details were. However, we would like to state that for this similar type of threat (malware infection), all of the groups regarded the same types of threat actors as the most likely ones. As shown by Holm et al. [78], there can be high degrees of uncertainty in data quality when expert judgement is used. Their experiments showed a significant negative correlation and a strong positive correlation between experience and calibration, suggesting that additional years' experience can both decrease and increase the calibration. It was outside the scope of our assessments to use calibration as the groups had different scope. However, the same facilitator was used in all workshops, and it became evident that the more experience he gained, the more effective the facilitation of the sessions. This is by no means a unique observation, but a lesson learned is that it may be useful to conduct a couple of pre-tests before the actual sessions.

To maintain confirmability, that is to reflect the voice of the participants from the evaluation, we have included representative statements in Section 5, as raw as possible. Though there is a translation bias from the Norwegian to the English language, we do not consider this to be of any significance. The recorded observational data and process notes have more of a subjective nature, but were shared with the participants after the sessions to allow for comments and show transparency.

Finally, we have to acknowledge that we are dealing with models about the future, where there can be rapid changes in the threat environment and unknown unknowns that no security expert or domain specialist can be expected to foresee. We find that the famous quote from Box and Draper [79] sums this up in an excellent way: "Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind".

7. Conclusion and further work

The threat likelihood approach has been developed to support security decision-making for storyless systems. It combines a number of existing concepts and techniques from risk management literature, expert judgements, and domain specific information in a systematic way. The main goal has been to create something applicable for real-life projects, efficient in terms of resource usage, and adjusted to what is the best data available. Through a systematic evaluation within a maritime case study, we have been able to assess the appropriateness of our contribution. The features supporting identification and quantification of threat actors, means, opportunity and motivation were all considered to provide some, strong, very strong or full support from representative groups in the cyber security and maritime community. Just as important as the threat likelihood value itself, is the ability to provide traceability on how the participants estimated it. Furthermore, in cases of underlying uncertainties, it was considered valuable to flag indication of this.

As for further work, it remains to develop better evidence on the generalisation of the results, both in terms of transferability to similar projects within the maritime context and also to different settings. This could be done using a similar research method for direct comparison, or through triangulation, mixing in quantitative methods applied to a larger set of projects and participants. The approach itself should be considered domain-independent, but the template should be adjusted to other contexts, e.g., critical systems related to water supply, energy, hospitals, and aviation to name a few. This requires a systematic gathering of relevant domain knowledge that is relevant and reusable for the threat assessments.

CRediT authorship contribution statement

Per Håkon Meland: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Dag Atle Nesheim:** Validation, Investigation, Resources, Writing – review & editing, Project administration, Funding acquisition. **Karin Bernsmed:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition. **Guttorm Sindre:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been partially supported by the projects "Cyber Security in Merchant Shipping - Service Evolution", funded by the Research Council of Norway with contract number 295969, and "CyberSec4Europe", funded by the European Union under the H2020 Programme Grant Agreement No. 830929.

References

[1] Franco EG, Kuritzky M, Lukacs R, Zahidi S. The global risks report 2021. Tech. rep., 16th Edition. World Economic Forum; 2021, URL http://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2021.pdf.

[2] ENISA. [ENISA] threat landscape 2020: Cyber attacks becoming more sophisticated, targeted, widespread and undetected. 2020, URL https://www.enisa.europa.eu/news/enisa-news/enisa-threat-landscape-2020.

[3] Burt T. Microsoft digital defense report. Tech. rep., Microsoft; 2020, URL https://www.microsoft.com/en-us/security/business/security-intelligence-report.

[4] Jalali MS, Siegel M, Madnick S. Decision-making and biases in cybersecurity capability development: Evidence from a simulation game experiment. J Strateg Inf Syst 2019;28(1):66–82. http://dx.doi.org/10.1016/j.jsis.2018.09.003, URL https://www.sciencedirect.com/science/article/pii/S0963868717304353.

[5] ISO. [ISO/IEC 27000:2018] Information technology - Security techniques - Information security management systems - Overview and vocabulary. Standard, International Organization for Standardization; 2018, URL https://www.iso.org/standard/73906.html.

[6] Schneier B. Threat modeling and risk assessment. In: E-Privacy. Springer; 2000, p. 214–29. http://dx.doi.org/10.1007/978-3-322-89183-9_20.

[7] Braiterman Z, Shostack A, Marcil J, Vries Sd, Michlin I, Wuyts K, et al. Threat modeling manifesto. 2020, URL https://www.threatmodelingmanifesto.org/.

[8] ISO. [ISO/IEC 27005:2018] Information technology - security techniques - information security management systems - information security risk management. Standard, International Organization for Standardization; 2018, URL https://www.iso.org/standard/75281.html.

[9] ISO. [ISO 31000:2018] Risk management guidelines. Standard, International Organization for Standardization; 2018, URL https://www.iso.org/iso-31000-risk-management.html.

[10] Stoneburner G, Goguen A, Feringa A. Risk management guide for information technology systems. NIST Spec Publ 2002;800(30).

[11] NIST. Cybersecurity framework version 1.0. 2014, URL https://www.nist.gov/cyberframework.

[12] Böhme R, Laube S, Riek M. A fundamental approach to cyber risk analysis. Variance 2019;12(2):161–85.

[13] Ahrend JM, Jirotka M. Anticipation in cyber-security. In: Handbook of anticipation: Theoretical and applied aspects of the use of future in decision making. Cham: Springer International Publishing; 2017, p. 1–28. http://dx.doi.org/10.1007/978-3-319-31737-3_26-1.

[14] Almukaynizi M, Marin E, Shah M, Nunes E, Simari GI, Shakarian P. A logic programming approach to predict enterprise-targeted cyberattacks. In: Data science in cybersecurity and cyberthreat intelligence. Cham: Springer International Publishing; 2020, p. 13–32. http://dx.doi.org/10.1007/978-3-030-38788-4_2.

[15] Hubbard DW, Seiersen R. How to measure anything in cybersecurity risk. Wiley Online Library; 2016.

[16] Santini P, Gottardi G, Baldi M, Chiaraluce F. A data-driven approach to cyber risk assessment. Secur Commun Netw 2019;2019. http://dx.doi.org/10.1155/2019/6716918.

[17] Tubío Figueira P, López Bravo C, Rivas López JL. Improving information security risk analysis by including threat-occurrence predictive models. Comput Secur 2020;88:101609. http://dx.doi.org/10.1016/j.cose.2019.101609, URL http://www.sciencedirect.com/science/article/pii/S0167404819301592.

[18] Kissoon T. Optimum spending on cybersecurity measures: Part II. J Inf Secur 2021;12(1):137–61. http://dx.doi.org/10.4236/jis.2021.121007.

[19] Al-Hadhrami N, Collinson M, Oren N. Modelling security risk scenarios using subjective attack trees. Risks secur internet syst 2021;2021. http://dx.doi.org/10.1007/978-3-030-68887-5_12.

[20] Brantly AF. Risk and uncertainty can be analyzed in cyberspace. J Cybersecur 2021;7(1). http://dx.doi.org/10.1093/cybsec/tyab001, tyab001.

[21] Buldas A, Laud P, Priisalu J, Saarepera M, Willemson J. Rational choice of security measures via multi-parameter attack trees. In: International workshop on critical information infrastructures security. Springer; 2006, p. 235–48. http://dx.doi.org/10.1007/11962977_19.

[22] Knez C, Llansó T, Pearson D, Schonfeld T, Sotzen K. Lessons learned from applying cyber risk management and survivability concepts to a space mission. In: 2016 IEEE aerospace conference. IEEE; 2016, p. 1–8. http://dx.doi.org/10.1109/AERO.2016.7500812.

[23] Llansó T, McNeil M, Pearson D, Moore G. BluGen: An analytic framework for mission-cyber risk assessment and mitigation recommendation. In: Proceedings of the 50th Hawaii international conference on system sciences. 2017.

[24] McNeil M, Llansó T, Pearson D. Application of capability-based cyber risk assessment methodology to a space system. In: Proceedings of the 5th annual symposium and bootcamp on hot topics in the science of security. 2018, p. 1–10. http://dx.doi.org/10.1145/3190619.3190644.

[25] ter Beek MH, Legay A, Lafuente AL, Vandin A. Quantitative security risk modeling and analysis with RisQFLan. 2021, arXiv preprint arXiv:2101.08677.

[26] Bagnato A, Kordy B, Meland PH, Schweitzer P. Attribute decoration of attack–defense trees. Int J Secur Softw Eng (IJSSE) 2012;3(2):1–35. http://dx.doi.org/10.4018/jsse.2012040101.

[27] Hong JB, Kim DS, Chung C-J, Huang D. A survey on the usability and practical applications of graphical security models. Comp Sci Rev 2017;26:1–16. http://dx.doi.org/10.1016/j.cosrev.2017.09.001.

[28] Paté-Cornell M-E, Kuypers M, Smith M, Keller P. Cyber risk management for critical infrastructure: a risk analysis model and three case studies. Risk Anal 2018;38(2):226–41. http://dx.doi.org/10.1111/risa.12844.

[29] Buldas A, Gadyatskaya O, Lenin A, Mauw S, Trujillo-Rasua R. Attribute evaluation on attack trees with incomplete information. Comput Secur 2020;88:101630. http://dx.doi.org/10.1016/j.cose.2019.101630, URL https://www.sciencedirect.com/science/article/pii/S0167404819301774.

[30] Mraković I, Vojinović R. Maritime cyber security analysis–how to reduce threats? Trans Marit Sci 2019;8(01):132–9. http://dx.doi.org/10.7225/toms.v08.n01.013.

[31] Tam K, Jones K. Macra: a model-based framework for maritime cyber-risk assessment. WMU J Marit Aff 2019;18(1):129–63. http://dx.doi.org/10.1007/s13437-019-00162-2.

[32] Kessler GC, Craiger JP, Haass JC. A taxonomy framework for maritime cybersecurity: A demonstration using the automatic identification system. TransNav: Int J Mar Navig Saf Sea Transp 2018;12(3):429–37. http://dx.doi.org/10.12716/1001.12.03.01.

[33] Svilicic B, Kamahara J, Rooks M, Yano Y. Maritime cyber risk management: An experimental ship assessment. J Navig 2019;72(5):1108–20. http://dx.doi.org/10.1017/S0373463318001157.

[34] You B, Zhang Y, Cheng L-C. Review on cyber security risk assessment and evaluation and their approaches on maritime transportation. In: Proceedings of the 30th annual conference of international chinese transportation professionals association. Houston, TX, USA; 2017. p. 19–21.

[35] Goldkuhl G. The research practice of practice research: theorizing and situational inquiry. Syst Signs Actions 2011;5(1):7–29, URL https://www.diva-portal.org/smash/get/diva2:480214/FULLTEXT01.pdf.

[36] Kitchenham BA. Evaluating software engineering methods and tool part 1: The evaluation context and evaluation methods. ACM SIGSOFT Softw Eng Notes 1996;21(1):11–4. http://dx.doi.org/10.1145/381790.381795.

[37] Kitchenham B. DESMET: A method for evaluating software engineering methods and tools. Tech. rep., University of Keele; 1996, URL http://www.osel.co.uk/papers/desmet.pdf.

[38] Zelkowitz MV, Wallace D. Validating the benefit of new software technology. Softw Qual Pract 1998;1(1). URL http://www.cs.umd.edu/~mvz/pub/sqp.pdf.

[39] Marshall C. Tool support for systematic reviews in software engineering. [Ph.D. thesis], University of Keele; 2016, URL https://eprints.keele.ac.uk/2431/1/MarshallPhD2016.pdf.

[40] Haga K, Meland PH, Sindre G. Breaking the cyber kill chain by modelling resource costs. In: International workshop on graphical models for security. Springer; 2020, p. 111–26. http://dx.doi.org/10.1007/978-3-030-62230-5_6.

[41] Robson C, McCartan K. Real world research. John Wiley & Sons; 2016.

[42] DHS. National maritime domain awareness plan for national strategy for maritime security. Tech. rep., Homeland Security Digital Library; 2013, URL https://www.hsdl.org/c/national-maritime-domain-awareness-plan-for-the-national-strategy-for-maritime-security/.

[43] Kontovas CA, Psaraftis HN. Formal safety assessment: a critical review. Mar Technol 2009;46(1):45.

[44] Cimpean D, Meire J, Bouckaert V, Vande Casteele S, Pelle A, Hellebooge L. Analysis of cyber security aspects in the maritime sector. Tech. rep., ENISA; 2011, URL https://www.enisa.europa.eu/publications/cyber-security-aspects-in-the-maritime-sector-1/at_download/fullReport.

[45] Caprolu M, Di Pietro R, Raponi S, Sciancalepore S, Tedeschi P. Vessels cybersecurity: Issues, challenges, and the road ahead. IEEE Commun Mag 2020;58(6):90–6. http://dx.doi.org/10.1109/MCOM.001.1900632.

[46] Chang C, Wenming S, Wei Z, Changki P, Kontovas C. Evaluating cybersecurity risks in the maritime industry: a literature review. In: Proceedings of the international association of maritime universities (IAMU) conference. 2019.

[47] Jacq O, Boudvin X, Brosset D, Kermarrec Y, Simonin J. Detecting and hunting cyberthreats in a maritime environment: Specification and experimentation of a maritime cybersecurity operations centre. In: 2018 2nd cyber security in networking conference. IEEE; 2018, p. 1–8. http://dx.doi.org/10.1109/CSNET.2018.8602669.

[48] CySiMS. Cyber security in merchant shipping. 2021, URL http://cysims.no/.

[49] IALA. VDES - VHF Data exchange system. 2020, URL https://www.iala-aism.org/technical/connectivity/vdes-vhf-data-exchange-system/.

[50] IEC. [IEC 61162-450:2018] maritime navigation and radiocommunication equipment and systems - digital interfaces - Part 450: Multiple talkers and multiple listeners - ethernet interconnection. Standard, International Electrotechnical Commission; 2018, URL https://webstore.iec.ch/publication/28704s.

[51] Meland PH, Bernsmed K, Wille E, Rødseth ØJ, Nesheim DA. A retrospective analysis of maritime cyber security incidents. In: Proceedings of the 14th international conference on marine navigation and safety of sea transportation. 2021.

[52] Williams J. OWASP risk rating methodology. 2020, [Online]. URL https://owasp.org/www-community/OWASP_Risk_Rating_Methodology.

[53] Pendse SG. Ethical hazards: A motive, means, and opportunity approach to curbing corporate unethical behavior. J Bus Ethics 2012;107(3):265–79. http://dx.doi.org/10.1007/s10551-011-1037-0.

[54] Van Ruitenbeek E, Keefe K, Sanders WH, Muehrcke C. Characterizing the behavior of cyber adversaries: The means, motive, and opportunity of cyberattacks. In: 40th annual ieee/ifip international conference on dependable systems and networks supplemental. 2010, p. 17–8, URL https://www.perform.illinois.edu/Papers/USAN_papers/10VAN01.pdf.

[55] Warikoo A. Proposed methodology for cyber criminal profiling. Inf Secur J Glob Perspect 2014;23(4–6):172–8. http://dx.doi.org/10.1080/19393555.2014.931491.

[56] Shinder DL, Cross M. Scene of the cybercrime. Elsevier; 2008, http://dx.doi.org/10.1016/B978-1-59749-276-8.X0001-5.

[57] Seafarer's professions and ranks. 2020, URL https://en.wikipedia.org/wiki/Seafarer%27s_professions_and_ranks.

[58] International ship and port facility security code. 2020, URL https://en.wikipedia.org/wiki/International_Ship_and_Port_Facility_Security_Code.

[59] Marine surveyor. 2020, URL https://en.wikipedia.org/wiki/Marine_surveyor.

[60] Dubay D. Why we will never see fully autonomous commercial ships. 2019, URL https://www.maritime-executive.com/editorials/why-we-will-never-see-fully-autonomous-commercial-ships.

[61] What is a data controller or a data processor? 2020, EU. URL https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/controller-processor/what-data-controller-or-data-processor_en.

[62] McKendall MA, Wagner III JA. Motive, opportunity, choice, and corporate illegality. Organ Sci 1997;8(6):624–47. http://dx.doi.org/10.1287/orsc.8.6.624.

[63] Hutchins EM, Cloppert MJ, Amin RM. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. In: Leading issues in information warfare & security research. Tech. rep., Lockheed Martin Corporation; 2010, URL https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/LM-White-Paper-Intel-Driven-Defense.pdf.

[64] Hutchins EM. The cyber kill chain. 2021, [Online]. URL https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html.

[65] Schneier B. Attack trees. Dr Dobb's J 1999;24(12):21–9.

[66] Meland PH, Sindre G. Cyber attacks for sale. In: 2019 international conference on computational science and computational intelligence. IEEE; 2019, p. 54–9. http://dx.doi.org/10.1109/CSCI49370.2019.00016.

[67] Calleja A, Tapiador J, Caballero J. A look into 30 years of malware development from a software metrics perspective. In: Monrose F, Dacier M, Blanc G, Garcia-Alfaro J, editors. Research in attacks, intrusions, and defenses. Cham: Springer International Publishing; 2016, p. 325–45. http://dx.doi.org/10.1007/978-3-319-45719-2_15.

[68] Casey T. Understanding cyber threat motivations to improve defense. Intel White Pap 2015. URL https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/understanding-cyberthreat-motivations-to-improve-defense-paper.pdf.

[69] Webster DA. Is there a difference between intent and motive?. The Law Office of David A. Webster, P.A.; 2019, [Online]. URL https://www.thewebsterlawoffice.com/blog/2019/june/is-there-a-difference-between-intent-and-motive-/.

[70] Casey T. Threat agent library helps identify information security risks. Intel White Pap 2007;2.

[71] Cockshott JE. Probability bow-ties: a transparent risk management tool. Process Saf Environ Protect 2005;83(4):307–16. http://dx.doi.org/10.1205/psep.04380.

[72] Meland PH, Bernsmed K, Frøystad C, Li J, Sindre G. An experimental evaluation of bow-tie analysis for security. Inf Comput Secur 2019;27(4):536–61. http://dx.doi.org/10.1108/ICS-11-2018-0132.

[73] Aust J, Pons D. A systematic methodology for developing bowtie in risk assessment: Application to borescope inspection. Aerospace 2020;7(7):86. http://dx.doi.org/10.3390/aerospace7070086.

[74] Bernsmed K, Frøystad C, Meland PH, Nesheim DA, Rødseth ØJ. Visualizing cyber security risks with bow-tie diagrams. In: International workshop on graphical models for security. Springer; 2017, p. 38–56. http://dx.doi.org/10.1007/978-3-319-74860-3_3.

[75] Meland PH, Bernsmed K. CySiMS threat likelihood approach template. 2021, http://dx.doi.org/10.5281/zenodo.4899525.

[76] Cruzes DS, ben Othmane L. Threats to validity in empirical software security research. In: Empirical research for software security. CRC Press; 2017, p. 275–300.

[77] Pols P. The unified kill chain: Designing a unified kill chain for analyzing, comparing and defending against cyber attacks. Tech. rep., Cyber Security Academy; 2017, URL https://www.unifiedkillchain.com/assets/The-Unified-Kill-Chain.pdf.

[78] Holm H, Sommestad T, Ekstedt M, Honeth N. Indicators of expert judgement and their significance: an empirical investigation in the area of cyber security. Expert Syst 2014;31(4):299–318. http://dx.doi.org/10.1111/exsy.12039.

[79] Box GE, Draper NR. Empirical model-building and response surfaces, Vol. 424. Wiley New York; 1987.