Kristian Haga

Breaking the Cyber Kill Chain by Modelling Resource Costs

Master's thesis in Computer Science
Supervisor: Per Håkon Meland and Guttorm Sindre
June 2020

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

SINTEF

Kristian Haga

Breaking the Cyber Kill Chain by Modelling Resource Costs

Master's thesis in Computer Science
Supervisor: Per Håkon Meland and Guttorm Sindre
June 2020

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Summary

The thesis presents the Resource Cost Model (RCM), a modelling approach to estimate the costs to an attacker to launch a cyberattack. The cost is estimated from the required resources of the attack. RCM combines the stagewise *Cyber Kill Chain* and attack trees to associate each resource with a phase in the cyberattack. The key concept of RCM is that to launch an attack the adversary must acquire all resources at each stage in the kill chain. If the defending party is able to deny the attacker access to a single resource, the kill chain is broken and the entire attack is mitigated. Further, from the properties of the identified required resources and the derived estimated cost, RCM suggests a set of probable attacker profiles through its resource based cybercriminal profiling methodology.

Kristian Haga has conducted the research supervised by Professor Guttorm Sindre and Senior Researcher Per Håkon Meland. Lead supervisor Sindre has monitored the research process, while co-supervisor Meland has provided feedback and advice.

A short paper based on work from this thesis was peer-reviewed and accepted for ”The Seventh International Workshop on Graphical Models for Security (June 22, 2020)” (GraMSec, 2020) and it will be published in Lecture Notes in Computer Science series by Springer. A pre-print of the short paper can be viewed in Appendix A.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The continuous digitization of all industries, infrastructure and society as a whole implies an immediate increased exposure to cybercrime. To combat this growing form of criminality, a clearer understanding of the costs, benefits and attractiveness of cyberattacks is necessary (Kshetri, 2006). This is in accordance with *Routine Activities Theory* introduced by Cohoen and Felson (1979). Routine Activities Theory is a crime rationale analysis stating which conditions that must be met for a rational crime to occur. Ekblom and Tiley (2000) and Grabosky (2001) extend Routine Activities Theory to include cybercrime and state that cybercrime will occur when the following four conditions are met: There exist an (1) accessible and attractive target, (2) the absence of a capable guardian and the presence of (3) a motivated offender with (4) the resources required to commit the crime.

The cyberattack on the private equity company Norfund and the ransomware attack on Hydro provide examples for accessible and attractive targets. Norfund reported on May 13th 2020 that they had been exposed to a serious case of fraud. The fraud was driven by a data breach giving the attackers access to a loan of USD $10 million (Norfund, 2020). The ransomware attack on Hydro hit its factories on March 19th 2019 and caused a full halt in production and the total cost is estimated to be in the range 550-650 million NOK (Hydro, 2019).

In the "Internet Organized Crime Threat Assessment 2019" (Europol, 2019) Europol states a persistent cyberthreat and that continuous efforts are needed to further synergize the network and information security sector and the cyber law enforcement authorities to improve the overall cyber resilience and cybersecurity. This describes the absence of a capable guardian.

Buldas et al. (2006) define the *Rational Attacker's Paradigm* to be that (1) there will be no attack if the attack is unprofitable and (2) the attacker chooses the most profitable way of attacking. From this it can be argued that as long as there exist profitable cyberattacks there will be motivated offenders.

For the latter condition, the required resources is not just a question of technical skills, but also a requirement that the offender is able to invest in software development and hardware acquisition, as well as the time it takes to plan, prepare and perform the attack.

Manky (2013) states that today's cybercrime exists in a highly organized form with marketplaces where agents possessing the technical skills required for an attack interacts with potential buyers of cybercrime. This marked enables any motivated offender with sufficient economical resources to utilize cyberattacks as a mean to commit crime. The fact that economical resources and not technical capabilities constrains the use of cyberspace to commit crime, raises the need to determine the cost of cyberattacks in order to determine the actual agents of attack. By determining the cost of a cyberattack the study derives the required resources, i.e. economical funds, for any malicious actor to commit a cybercrime.

From the above we formalize the research questions that explore the possibility of modelling a cyberattack based on its required resources in order to derive rational agents of attack:

1. How can the cost of launching a cyberattack be estimated?

2. Which rational agents of attack do the required resources and the cost of a cyberattack imply?

Determining the cost of a given cyberattack in order to derive possible malicious actors presents an improved analysis of cyberthreats. Not limiting the scope of attackers based on their technical capacities, and rather focus on the cost of an attack, enable defending actors to address a broader specter of possible adversaries. The derived set of possible malicious agents will include agents that lack technical capacities, but utilize the cybercrime-as-a-service marked.

This study presents the *Resource Cost Model* (RCM) and its installation the *Interactive Resource Cost Model* (IRCM). RCM is a modelling approach that maps resource costs to each stage of a cyberattack, and derives the total costs of the attack. The model utilizes principles from Schneier's (Schneier, 1999) attack trees and the Lockheed Martin's cyber kill chain introduced by Hutchins et al. (2011), both already widely known in the security community, to structure this approach. IRCM has been developed as a dedicated prototype tool to simplify and visualize the modelling process, and the study has completed the first rounds of iterative evaluation among experts. The installation is validated and instantiated in a maritime context, but it is thought to be a generic tool. The IRCM tool is able to interactively show calculations and extract potential offenders based on a built-in library from available cybercriminal profile literature. The goal of RCM and its installation IRCM is to improve the accuracy of threat analysis, and especially increase the understanding and awareness of cyberthreats among sectorial domain stakeholders. RCM and IRCM are instantiated and validated in a maritime context, but both are thought to be generic modelling tools for all sectors and domains.

Wortman and Chandy (2020) describe the security risk as an equation of probability, impact and cost of attack. The cost of a cyberattack is expressed as the most difficult value to verify, in addition to the monetary value providing a method of comparison. Shang et al. (2019) propose an information security risk assessment method based on a attack tree model with fuzzy theory and probability risk assessment technology, which is applied in a risk scenario of a ship control system. Boyes and Isbell (2017) present a code of practice for cybersecurity in maritime by presenting a cyber security assessment followed by guidelines on how to develop a cybersecurity plan. Further, Boyes and Isbell (2017) present threat actor personas in maritime. Tam and Jones (2019) propose MaCRA a model-based framework for maritime cyber-risk assessment using three main criertia: (1) system vulnerability and effect, (2) ease-of-exploit, and (3) reward of a successful attack. The three latter studies investigate how to model a cyberattack in maritime and mitigation efforts, but do not account for the cost of an attack.

The study is structured as follows. The method of research is described in chapter 2. Chapter 3 gives a brief background on cybersecurity, cybercrime-as-a-service and crimeware markets, criminal behavior models, and cyberattack modelling approaches. Following, chapter 4 presents a literature review on cybercriminal profiling. Chapter 5 presents the *Resource Cost Model* (RCM) and chapter 6 presents the derived methodology for resource based cybercriminal profiling. Further, chapter 7 describe how the study evaluated and validated RCM and its installation IRCM. Chapter 8 discusses the results of the evaluation and identifies further work. Finally, chapter 9 concludes to what extent the thesis answer the research questions presented above.

Chapter 2

Method

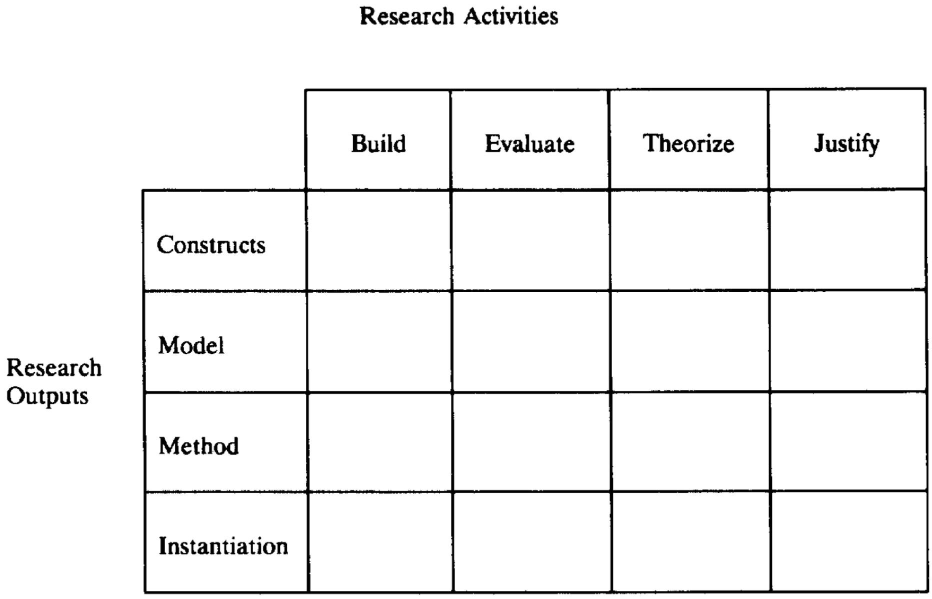
This study incorporates a design and creation research strategy, building information system artifacts.

Using the two dimensional framework driven by the distinction between research outputs and research activities in IT research proposed by March and Smith (1995), see Figure 2.1, the research has developed a modelling approach that estimates the cost of carrying out a cyberattack and from this cost derives a set of probable attacker profiles. To validate and justify the modelling approach and provide a proof-of-concept, the study has built an interactive installation of the model in the form of a web application called *Interactive Resource-Cost Model (IRCM)*. The interactive model allows users to model cyberattacks of their choosing, while concurrently deriving the total cost of the attack and probable attacker profiles.

The interactive installation of the model serves as a vehicle for evaluating cyberattacks from both a technical and an economical perspective. Through evaluating the cost of the required resources in order to launch an attack, the study learns which personas that are probable rational agents of attack. This places the research in the Information System research domain.

The research finds it self in the methodological pragmatism research paradigm. Goldkuhl (2012) describes the paradigm as being concerned with how knowledge is created and emphasizes the active role of the researcher in creating data and theories. Here, how knowledge about a rational adversary is created through deriving the cost of a cyberattack.

As the study seeks to mitigate the cyberthreat through the Interactive Resource-Cost Model, it applies a design science methodology. Hevner and Chatterjee (2010) states that design science supports a pragmatic research paradigm that calls for the creation of innovative artifacts to solve real-world problems. Further, a design science research improves the effectiveness and utility of the IT artifacts in the context of solving a real-world business problem. To improve the effectiveness and utility of IRCM, the study incorporated the *Design Thinking* framework. Design thinking defines an iterative workflow of research activities throughout the development of IT artifacts.



Research Activities

		Build	Evaluate	Theorize	Justify
Research Outputs	Constructs				
	Model				
	Method				
	Instantiation				

Figure 2.1: Research framework for IT research proposed by March and Smith (1995)

2.1 Design Thinking

The principles of design thinking were first described in Simon (1996) as an iterative process which seeks to understand the user, challenge assumptions and redefine problems in an attempt to identify solutions. Brown and Katz (2011) sums up design thinking as a problem solving approach, crystallized in the field of design, which combines a user-centered perspective with analytical research. Design thinking is a tool to find the intersection of technical feasibility, economic viability, and desirability by the user through an experimental, user test driven process.

With the user, businesses aiming to mitigate the cyberthreat, at the center of the model development, the study ensures that the Interactive Resource-Cost Model conveys the threat of a cyberattack and its potential rational attackers to the actors able to mitigate the threat or investigate an occurred attack.

The study incorporates the five-phase workflow proposed by the Hasso-Plattner Institute of Design at Stanford (Hasso Plattner Institute of Design, 2020a) (Hasso Plattner Institute of Design, 2020b) (Soegaard, 2018). The five stages are:

- *Empathize* - with your users
- *Define* - your users needs, their problem and your insights
- *Ideate* - by challenging assumptions and creating ideas for innovative solutions
- *Prototype* - to start creating solutions
- *Test* - to validate and evaluate the solution

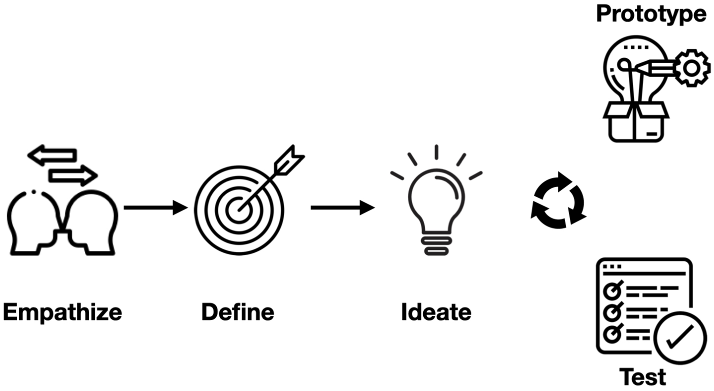


Figure 2.2: Hasso-Plattner Institute Design at Stanford (Hasso Plattner Institute of Design, 2020a) (Hasso Plattner Institute of Design, 2020b) design thinking workflow

The study followed the workflow in Figure 2.2; firstly completing the *Empathize* stage followed by the *Define* stage, before entering an iterative process through an Ideate-Prototype-Test-loop. The result of the iterative process is the creation of the Interactive Resource-Cost Model, which features and functions are anchored in insights derived through the *Empathize* and *Define* stage. During the Ideate-Prototype-Test-loop, the study revisited the *Define* stage when observations suggested to tweak the problem definitions.

2.1.1 Empathize

The *Empathize* stage of the design thinking process seeks to accomplish an empathetic understanding of the problem: The increasing cyberthreat. This involves to research the topic in depth and engaging and empathizing with actors within the problem domain.

As IRCM is instantiated in a maritime context, the study reviewed the literature on the maritime cybersecurity domain in order to understand the cyberthreat. This review is presented in the preliminary work (Haga, 2019). Further, the research conducted a literature review, presented in chapter 4, on cybercriminal profiling. This served the purpose of gaining knowledge on cybercriminal profiling methods and profiles, in order to empathize with the adversary. It can be argued that an understanding of the adversary, enhance the accuracy of the study's model deriving rational adversaries based on the total cost of a cyberattack.

Later, equipped with IRCM, the study consulted the maritime and cybersecurity sector on their requirements for an interactive tool to present the cyberthreat on maritime and derive malicious actors. This put the user back at the center of development after researching the topic in general.

2.1.2 Define

Through the *Define* stage the information gathered in the former *Empathize* stage is utilized to define the research questions of the thesis - the problem that the Interactive Resource-Cost Model aims to solve.

The research questions were synthesized based on the literature reviews conducted in the first part of the *Empathize* stage. These questions present a formal, academic formulation of the new knowledge the study aims to add to the domain.

By consulting the maritime sector on their requirements of an interactive tool, the research questions were revised into a user-centered formulation.

1. How can an interactive modelling tool aid actors in maritime to estimate the costs of a cyberattack?

2. Which rational agents of attack must actors in maritime be aware of based on the costs of a maritime cyberattack?

This aids to establish features and functions required by IRCM and how such a tool achieves to answer the research questions through user interaction.

2.1.3 Ideate

The *Ideate* stage consists of generating ideas to solve the problems defined in the previous stage, based on insights from the *Empathize* stage.

The Resource-Cost Model, presented in chapter 5, is the fundamental idea to the solution and served as the platform from which the research generated ideas of features and functions to the Interactive Resource-Cost Model in later iterations.

The *Ideate* stage was revisited after each iteration through the following *Prototype* and *Test* stages to tune and revise existing ideas and generate new ideas aiming to improve the information system artifacts at the center of the study.

In the second iteration of the design thinking cycle, the research decided during the *Ideate* stage on realizing the derived model from the first iteration as a web application. The decision fell on a web application, instead of a software or operative system native application, because it allows for a more rapid *Prototype* stage building a minimum viable product (MVP). Most well-established web application frameworks, e.g. React and Ruby on Rails, provide scaffolds for working out of the box web applications. A web application installation of the model is also easier to distribute than a software solution. This was vital in order to allow for remote user testing.

The study chose Ruby on Rails (RoR) as its web development framework. The pick of framework was based on RoR being well-established and its convention-over-configuration design.

RoR was released in 2005 and is today used by almost a million websites, including GitHub, Airbnb and Hulu (Nowak, 2020). From this it follows a rich set of easily accessible educational resources and a well reviewed documentation.

Frameworks that fall under the convention-over-configuration software design paradigm, attempts to decrease the number of decisions a developer must take without losing flexibility (Wikipedia, 2019c). In context of RoR this means that there is a ”right way” - the

”RoR-way” - of writing code. When following the conventions and coding in the ”RoR-way”, RoR facilitates time consuming tasks such as routing and database management which are required by any web application. This propels rapid, solid development and lets the developer focus on app features and logic.

In addition, RoR provides an extensive scaffold for a basic web application. The scaffold comes with a ”RoR-way” file structure including example files for views, controllers, scss-styling and more. This launches the development forward into the *Prototype* stage constructing a MVP.

Finally, the collaboration between RoR and the web hosting platform Heroku, makes the process of publishing a web application trivial. For us, this was important to rapidly begin remote user testing.

In addition to solving design issues identified during the *Test* stage of the second iteration, the third iteration through the *Ideate* stage resulted in the resource-based cybercriminal profiling methodology presented in chapter 6.

2.1.4 Prototype

In the *Prototype* stage the study iteratively implemented the new ideas and reviewed features from the previous *Ideate* stage, resulting in improved versions of the model.

In the first design thinking cycle the *Prototype* stage implemented the ”pen and paper” version of the model described in chapter 5. This ”pen and paper” version made it possible to validate the model in a maritime setting by modeling a known cyberattack. The validation is described in detail in Appendix B and served as a proof-of-concept for the model.

The second iteration of the stage produced a minimum viable product (MVP) of the Interactive Resource-Cost Model. Ries (2011) defines a MVP as the version of a new product which allows developers to collect the maximum amount of validated learning about customers with the least effort. An MVP implements the core features and functions of the product such that users can provide feedback at an early stage through product testing.

The MVP consisted of an information page presenting the key concepts of the IRCM and functionality enabling users to build Resource Cost Models with an arbitrary number of resources and resource alternatives. The MVP is presented in detail in section 7.1.

The third iteration of the *Prototype* stage implemented the cybercriminal profiling feature as well as tweaking flawed features and functionality identified during testing of the MVP. The resulting version of IRCM is presented in detail in section 7.2.4.

2.1.5 Test

The *Test* stage involves conducting user tests, which serve the purpose to revel flawed or confirm functioning solutions. The user test drives the experimental development process by redefining problems and informs the understanding of the users, the conditions of use, how people think, behave, feel and to empathize. Iterative user tests derive an understanding of the product and its users.

The user tests are presented in detail in chapter 7 and the results are discussed in chapter 8.

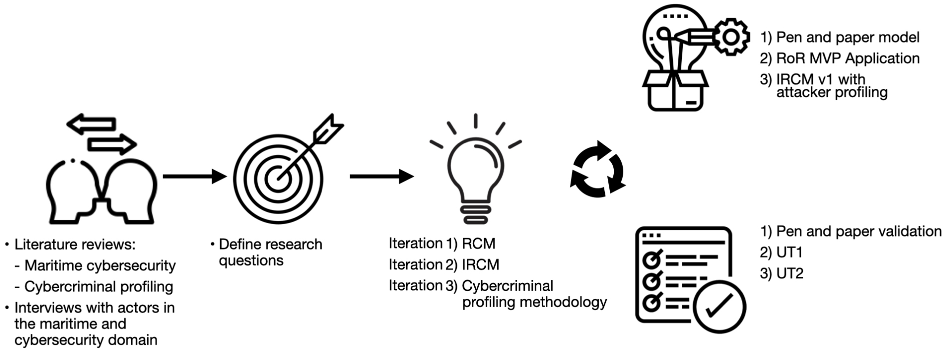

1) Pen and paper model
2) RoR MVP Application
3) IRCM v1 with attacker profiling
• Literature reviews:
- Maritime cybersecurity
- Cybercriminal profiling
• Interviews with actors in the maritime and cybersecurity domain
• Define research questions
Iteration 1) RCM
Iteration 2) IRCM
Iteration 3) Cybercriminal profiling methodology
1) Pen and paper validation
2) UT1
3) UT2

Figure 2.3: Main tasks in the study mapped to the stages and related iteration in the design thinking workflow

Ideally the study should have completed many iterations through the Ideate-Prototype-Test-loop with only minor changes in each iteration to amplify the experimental, user test driven research method. Unfortunately, the difficulty of performing testing with actors in the maritime and cybersecurity sector due to their time constraints limited the study to fewer iterations with larger improvements to the model in each iteration. The main tasks at each stage through each iterations are shown in Figure 2.3.

Chapter 3

Background

3.1 Cybersecurity

In the Committee on National Security Systems Instruction Glossary, (CNNSI, 2015), the National Institute of Standards and Technology (NIST) defines the terms cybersecurity, computer security or information technology security as the measures and controls provided in order to preserve the confidentiality, integrity and availability of the information processed and stored on a computer.

Confidentiality, integrity and availability are commonly referred to as the CIA-triad which states the elementary security measures of any digital information system.

CNNSI (2015) defines the terms in the CIA-triad:

- **Confidentiality** - The ability to preserve authorized restrictions on information access and disclosure, including protecting proprietary information and personal privacy.
- **Integrity** - To hinder improper information modification or destruction, and to ensure information authenticity and non-repudiation.
- **Availability** - An information system ability to ensure timely and reliable access to and use of information on the system.

3.1.1 Cyberattack

A cyberattack seeks to compromise at least one of the components in the CIA-triad. NIST defines in the CNNSI (2015) glossary a cyberattack to be an attack, via cyberspace, targeting the victim's use of cyberspace. A cyberattack has the goal of disrupting, disabling, destroying or maliciously control a computer environment or infrastructure. This include violating the integrity of data or stealing information.

The Internet Engineering Task Force (IEFT) in Shirey (2007) categorize a cyberattack as passive or active.

In a *passive cyberattack* the attacker seeks to make use of or disclose information on the target system, without altering any data or affecting the resources of the target system.

In contrast, in an *active attack* the adversary creates a false data stream or makes modifications to the existing data stream, with the purpose of alter system resources or affect their operation.

Hutchins et al. (2011) present the *Intrusion Kill Chain* as a tool for describing each stage of a cyberattack. The kill chain analysis illustrates how an attacker must successfully and consecutively progress through all stages in order to complete a cyberattack. The seven stages of the kill chain are:

1. *Reconnaissance* - Research, identification and selection of target.
2. *Weaponization* - Coupling a malware with a deliverable payload, e.g. image, PDF or Microsoft Office document.
3. *Delivery* - Transmission of the weapon to the targeted environment, e.g. an email attachment or USB-drive.
4. *Exploitation* - Triggers malicious code. Ranges from vulnerabilities or auto-executing features in host's operating system, to users triggering execution. Factors enabling/triggering the execution of malicious code.
5. *Installation* - Installation of malware or remote access.
6. *Command and Control (C2)* - C2 channel provides ”hands on the keyboard” access inside the target environment.
7. *Actions on Objectives* - Actions required to achieve the goal of attack.

Diogenes and Ozkaya (2018) elaborates on the *Reconnaissance* stage as the phase where attackers search for vulnerabilities they can leverage for an attack. This involves gathering data, identifying users of and loopholes in the target system. When reconnaissance is done right, the target is not aware that it has happened, thus only including *passive attacks*. Further, reconnaissance techniques can be classified as *External* - happening outside the target's network - or *Internal* - occurring within the target's systems. Internal reconnaissance is usually aided by software tools interacting with the target system. Below we present common reconnaissance techniques presented in Diogenes and Ozkaya (2018).

Dumpster Diving

Dumpster Diving describes attackers going through disposed of obsolete computers or external storage devices. Many organizations do not have a rigorous process for disposing digital devices and as a consequence the devices may expose sensitive information. Diogenes and Ozkaya (2018) state that obsolete storage devices may give an adversary information on the internal setup of an organization, passwords stored in browsers, user information and even elevated internal system access.

Social Media

As the future employer or secret admirer, cybercriminals go through social media accounts for victim information. Public available data on social media can be exploited to guess passwords or answer secret questions used to reset some accounts. Information interesting to an attacker is date of birth, parent's name, pet names, school names etc.

Further, social media facilitate for digital identity theft. Using public available information on the victim, a hacker can create a fake account bearing the identity of another person. Diogenes and Ozkaya (2018) describe how hackers track information on high-ranking, key personnel in an organization, then leveraging this information for favors or system access by impersonating the high-ranking employee through a fake social media account.

Social Engineering

Social engineering is developed to exploit human nature, thus no security system can shield itself from this type of threat. In social engineering malicious actors leverage that humans are sympathetic, trusting of friends, show-offs and obedient to higher authorities. Social engineering happens on the outside of the target's network, thus it is classified as external reconnaissance.

Phising - According to the *Common Attack Pattern Enumeration and Classification (CAPEC)* by MITRE (MITRE, 2019c) *Phising* is a social engineering technique where the attacker masquerades as a legitimate entity in order to prompt the victim to disclose information. Phising is often used to gather authentication credentials which again can be used in a *Privilege Escalation* attack. It is most common for phishing attacks to be carried out through spam emails containing malicious links harvesting data or compromising the system when clicked. In addition, Chapter 4 in Diogenes and Ozkaya (2018) present how a Facebook post was used for phising.

Note that pishing can also be used in the *Delivery* stage, as well as to be the main attack method used throughout all stages of the kill chain when compromising a system.

Pharming - MITRE states in MITRE (2019a) that a *Pharming* attack discloses information by fooling the victim into entering sensitive data into supposedly trusted locations, e.g. an online bank. Performing a pharming attack the adversary impersonates such legitimate sites in order to harvest sensitive information by users being under the impression of interacting with the legitimate site.

Pretexting - According to Hadnagy (2010), in *Pretexting* the attacker constructs an elaborate lie that is well-researched so as to appear as legitimate to the target. The lie is then leveraged to get the target to divulge information or perform actions favorable to the attacker. An example is attackers posing as the CFO of a company, requesting accountants for a payment to some fake project account controlled by the attacker. Further, attackers are known to impersonate tax officials, police officers, dept collectors and other actors viable to request a target for money.

Diversion theft - In *Diversion theft* the attacker cons delivery and transportation companies that their deliveries are requested elsewhere. This enables the attacker to gain access to a delivery and the possibility to install rootkits, mal- or spy-ware before posing as a legitimate courier delivering the flawed product to the target. By misguiding a delivery, an attacker can also impersonate a courier with a scheduled delivery while the legitimate courier is on a detour.

Water holing - *Water holing* takes advantage of the trust users put in websites they frequently visit, e.g. forums and online banking. Diogenes and Ozkaya (2018) claim that even careful and cyber-suspicious users are more likely to act carelessly on familiar sites. Hence, water holing is often used to target IT knowledgeable victims.

Further, water holing can be combined with *pharming* of familiar sites to the victim. This requires prior reconnaissance of the victim and a development effort by the attacker, thus implying that the target was deliberately chosen.

An example of a water holing attack can be to put up a post on stackoverflow.com containing a malicious link.

Baiting - *Baiting* involve planting an external storage device containing malicious code in a place where a curious, greedy and careless person in the target organization will stumble upon the device. Then the attacker simply waits for the device to be plugged into the target system and executes the malware.

Diogenes and Ozkaya (2018) claim that bating has a high success rate due to the greedy and curious nature of humans. We simply can't constrain our self from plugging in a memory stick with ”confidential” written in red ink on it.

Quid pro quo - This is the common low skill-level attack where the attacker call random numbers claiming to be from technical support and offers technical support. A careless victim following the guide of ”technical support” may give the attacker credentials, credit card numbers, system access etc.

Sniffing and Scanning

Conducting *internal* reconnaissance an adversary generally eavesdrops on traffic in the target's network according to Diogenes and Ozkaya (2018). Sniffing tools are designed to conduct *passive Man in the Middle* (MITM) attacks.

By MITRE (2019b) a MITM attack describes an attacker that places himself in the communication channel between two communicating parties. Whenever a party attempts to communicate, the data being sent is intercepted by the attacker before it is relayed to the intended receiving party. When the attacker has intercepted the data, it can observe and alter the data before it is relayed. Neither the sender nor the intended receiver is aware of the transparent interposition of the attacker. The supposedly most sophisticated and largest MITM attack was conducted over several years after the 9/11 terrorist attack in 2001 by NSA and other state surveillance agencies in the US according to whistleblower Edward Snowdon.

Available sniffing tools include Prismdump, tcpdump, Wireshark and Nmap.

Note that a MITM attack altering data will classify as an *active attack*.

The next six stages in the *Intrusion Kill Chain* involve compromising the target using the information obtained from the reconnaissance stage.

Below the study states some of the most common types and patterns of cyberattacks compromising targets according to:

- Allianz Risk Barometer (AGCS, 2020) - Survey with 2,718 respondents in 22 industry sectors from 102 countries.

- Mørketallundersøkelsen (NSR, 2018) - Survey by Næringslivets Sikkerhetsråd (NSR) with 1500 respondents, all Norwegian, in ten industry sectors.

- Diogenes and Ozkaya (2018)

Denial of Service

A Denial of Service (DoS) attack seeks to prevent access to or delay time-critical operations of the target resource, thus compromising the *availability* of the target according to CNNSI (2015) and Shirey (2007). Launching a DoS attack, the adversary floods the target with requests. The Cybersecurity and Infrastructure Security Agency describes in NCCIC (2018) a DoS attack as a success when the superfluous traffic overloads the target system such that legitimate requests fail.

Today, most DoS attacks leverage multiple sources of traffic to flood the target system. This method is coined Distributed Denial of Service (DDoS). Usually attackers leverage Internet of Things (IoT) devices, e.g. baby monitors, smart speakers etc., for a DDoS attack. IoT devices are attractive when creating a botnet used for a DDoS, because IoT devices are known to have poor security and the sheer number of IoT devices online.

A DDoS attack targeted Wikipedia on September 6 and 7, 2019, bringing the site down in Germany and other parts of Europe according to Wikipedia on Twitter (Wikipedia, 2019a).

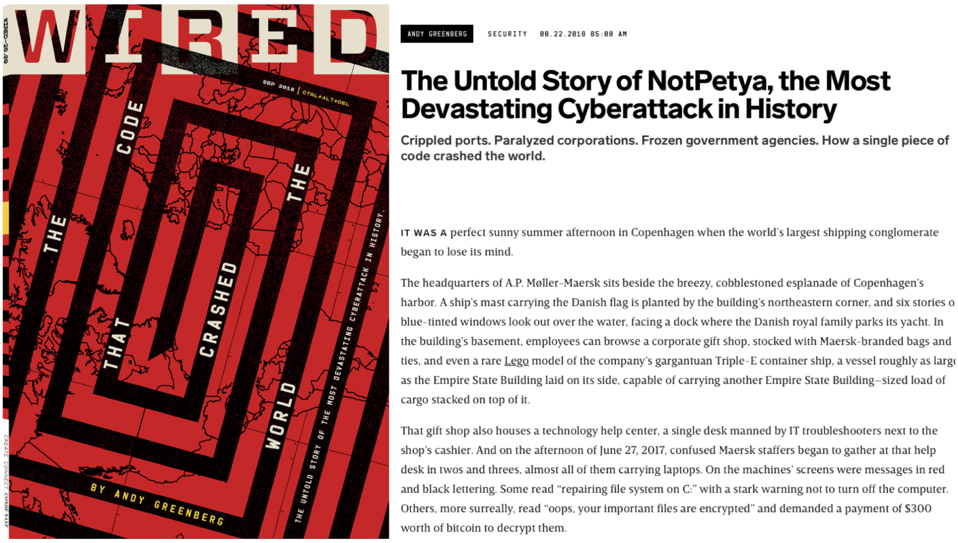
Code Execution

In a Code Execution (CE) attack the goal of the attacker is to execute arbitrary commands or code on the target machine or process according to Wikipedia (2019b). The execution of commands classifies such attacks as *active attacks*. CE violates the *integrity* and *confidentiality* of the target. In addition a CE attack may be leveraged to compromise the *availability* of the target.

The *EternalRomance* exploit is an example of vulnerabilities allowing for remote code execution - the ability to execute arbitrary commands on the target system over a network (Nahorney, 2019), (Arntz, 2018). EthernalRomance was leveraged by the NotPetya attack on the shipping line Møller-Mærsk in 2017. The attack caused the congestion of several ports world wide and business losses in excess of USD $250-$300 million as reported by Moller-Maersk (2017) and AGCS (2019).

Privilege Escalation

NIST states in Lee Badger (2016) that a Privilege Escalation (PE) attack is accomplished when the attacker achieves a higher privilege level than intended. The core of a PE attack, elevating ones privileges, classifies as a *passive attack*, but the unintended access may be leveraged in *active attacks*. A PE can be leveraged to compromise all three components of the CIA-triad by being the entry point of various malicious actions.


WIRED
WIRED–26.09
SEP 2018 | CTRL+ALT+DEL
THE CODE THAT CRASHED THE WORLD
THE UNTOLD STORY OF THE MOST DEVASTATING CYBERATTACK IN HISTORY.
BY ANDY GREENBERG
CREATE. CONNECT. CONDÉ NAST

ANDY GREENBERG SECURITY 08.22.2018 05:00 AM

The Untold Story of NotPetya, the Most Devastating Cyberattack in History

Crippled ports. Paralyzed corporations. Frozen government agencies. How a single piece of code crashed the world.

IT WAS A perfect sunny summer afternoon in Copenhagen when the world's largest shipping conglomerate began to lose its mind.

The headquarters of A.P. Møller-Maersk sits beside the breezy, cobblestoned esplanade of Copenhagen's harbor. A ship's mast carrying the Danish flag is planted by the building's northeastern corner, and six stories o blue-tinted windows look out over the water, facing a dock where the Danish royal family parks its yacht. In the building's basement, employees can browse a corporate gift shop, stocked with Maersk-branded bags and ties, and even a rare Lego model of the company's gargantuan Triple-E container ship, a vessel roughly as larg as the Empire State Building laid on its side, capable of carrying another Empire State Building–sized load of cargo stacked on top of it.

That gift shop also houses a technology help center, a single desk manned by IT troubleshooters next to the shop's cashier. And on the afternoon of June 27, 2017, confused Maersk staffers began to gather at that help desk in twos and threes, almost all of them carrying laptops. On the machines' screens were messages in red and black lettering. Some read "repairing file system on C:" with a stark warning not to turn off the computer. Others, more surreally, read "oops, your important files are encrypted" and demanded a payment of $300 worth of bitcoin to decrypt them.

Figure 3.1: Article from Wired (Greenberg, 2018) describing the NotPetya attack on Møller-Mærsk

The Mirai botnet which was used to launch a DDoS attack on the DNS provider Dyn in October 2016 leveraged at its core privilege escalation. Antonakakis et al. (2017) explain how the Mirai malware conducted the primitive strategy of trail and error using common passwords to log into, thus escalating privilege on a wast number of IoT devices. These devices created a botnet launching an attack resulting in the inaccessibility of several websites, including GitHub, Netflix and Twitter.

Spoofing

CNNSI (2015) defines Spoofing as an attack which seeks to induce a resource or user to take incorrect actions, thus classifying spoofing as an *active attack*. Spoofing is accomplished by the attacker faking the sending address of a transmission.

Spoofing attacks takes many suits and range from e-mail spoofing - spammers hiding their origin behind a legitimate address - to GPS Spoofing - altering the perceived location of a GPS system - as shown possible by Psiaki and Humphreys (2016).

Extortion attacks

Extortion attacks involve to hold computer files ransom or threaten to release damaging information regarding the victim. In both instances, the attacker demands to be paid in order to give back the files or withholding damaging information. As the offender only seeks to violate the confidentiality of the system and to hold data hostage, not altering it, extortion attacks classifies as *passive attacks*.

A common technique used to hold files ransom is the use of *Ransomware*. Liska and Gallo (2016) define ransomware as a class of malware that is used to digitally extort victims into paying a file release or access fee. The malware typically encrypts files on the victim's machine or makes the system unavailable by change of usage rights. Meland et al. (2020) claim that ransomware has democratised cybercrime through the *Ransomware-as-a-Service* market offering working out of the box ransomware, thus actors with little computer knowledge can leverage such an attack.

A notoriously famous ransomware attack is WannaCry in May 2017. The malware infected hundreds of thousands of computers in over 150 countries. The ransomware encrypted the files on the infected computer before asking for $300 to be paid in order for the malware to decrypt the files (Chen and Bridges, 2017).

In addition to the immediate, possible exposure of sensitive data following an information disclosure attack, the IOCTA 2018 report (Europol, 2018) by Europol presents the possibility of leveraging information disclosure combined with GDPR as an attack vector for ransom attacks. The report suggest the attacker might disclose user data and use this data as leverage to blackmail the victim. The victim might find it cheaper to pay the attacker to release its data, than to report the loss of user data resulting in a fine due to GDPR.

Diogenes and Ozkaya (2018) claim that the recent increase of the number of extortion attacks shows that these attacks are becoming preferred over trying to sell disclosed data. Data is often more valuable to its owner than to third parties.

Data manipulation attacks

In *data manipulation attacks* the adversary compromises the system through altering data, thus violating the *integrity* of the system in an *active* attack. An attacker can manipulate data with the intent to disrupt or sabotage the target's operations or cause the target to distrust the integrity their own data. This make data manipulation attacks attractive to competing companies. Further, an attacker can demand ransom to re-alter manipulated data. Data manipulation is often difficult to detect, thus it may have long lasting effect.

Diogenes and Ozkaya (2018) argue that data manipulation attacks will be the next stage of cybercrime and pose a severe threat on health care, financial and government data. An imminent, frightening scenario in context of today's, April 2020, COVID-19 pandemic, is the manipulation of data on infected citizens. This data and data on the movement of infected people is used to warn others through a smart phone application tracking the user's movement. By tagging uninfected citizens as infected, it is not hard to imagine the outbreak of chaos as other users, based on misinformation, fear they have interacted with infected people. This could put millions of citizens in unnecessary quarantine, further halting the economy.

Backdoors

Sparks et al. (2009) define a *Backdoor* as a method to covertly send and receive data from the system. Backdoors have been found to be planted and embedded in commercially shipped software.

The National Security Agency (NSA) has been accused of implementing such backdoors. Diogenes and Ozkaya (2018) claim that non state actors have started implementing their own backdoors by compromising the development of software.

Mobile device attacks

Symantec (2019) and Symantec (2017) show an increase of cyberattacks on mobile devices in the last five years. This is supported by the findings of Meland and Sindre (2019), showing that mobile device malware are among the most bought cybercrime tools on the Darkweb.

Common malware families reported by Symantec (2019) are ransomware and spyware stealing personal information. It is also reported that mobile devices are infected through users clicking fraudulent adverts, man-in-the-middle attacks and scripting attacks leveraging the low level of security on mobile web browsers. Further, Diogenes and Ozkaya (2018) point out that users tend to care less about securing their phone than laptop through antivirus software.

Hacking everyday IoT devices

Internet of Things (IoT) devices are non-conventional digital devices connected to the internet such as CCTV cameras, printers and the wast family of ”smart” devices, e.g. smart-speakers, -fridges, -TVs, -ovens, -reading lights etc.

A commonality of IoT devices is their lack in security features, thus making them easy targets. Many devices are only password protected and it has been shown through the Mirai attack that a significant number of the passwords are weak (Antonakakis et al., 2017).

Diogenes and Ozkaya (2018) confirm that printers and smart TVs have been hacked to spy on the target through relaying the files sent to the printer or using the smart TV to record and relay all audio. The latter was made possible on Samsung TVs by the ”Weeping Angel” exploit.

Cloud-hacking

Cloud services provide an on demand, scalable computing capacity to its clients. This ensures accessibility to the end-users, while the company hosting its service in the cloud only pay for the actual usage of computing power and doesn’t have to manage any local servers them self.

The security issue of the cloud rises from the fact that all users of the same cloud service share memory, network interface and CPUs. Hence, all an attacker has to do is to be able to traverse the data stored on a single cloud server in order to compromise all services hosted on that server. This leaves the cloud vendor with a major security responsibility, while it limits the security features an organization using the cloud can provide. The security of a user of a cloud service is never better than the security provided by the vendor.

Symantec (2019) states that the main reason for vulnerable clouds are poor configurations by users and hardware exploits such as Spectre and Meltdown.

Vulnerability exploitation

A system vulnerability is exploited when hackers take advantage of bugs in a software system. The bugs can origin within the operating system, the kernel or a web-based system. The bugs provide loopholes attackers exploit to conduct malicious acts. Loopholes unknown to the product provider before they are discovered to have been be used in a cyberattack are coined *Zero-day exploits*. Two common techniques to discover zero-day vulnerabilities are *Fuzzing* and *Source Code Analysis*.

Fuzzing is the process of recreating a system while searching for vulnerabilities. Through reconstructing the target system an attacker determines safety considerations made by the developers and which bugs they encountered. Although cumbersome, fuzzing provides an adversary a full understanding of the target system.

Source code analysis is possible on systems that release their source code. Attackers can either review the source code manually looking for vulnerabilities or use tools like IDA PRO to discover vulnerable code, enabling attack vectors such as *SQL injections* or *Cross-Site-Scripting*.

Today, many tech companies publish their code while offering large finder's fees for reported vulnerabilities discovered by the community. As an example, Microsoft offers bounties up to USD $300,000 for reported vulnerabilities in their cloud platform Microsoft Azure (Microsoft, 2020). These rewards are meant to encourage reporting over exploitation of zero-day exploits.

Due to the fact that zero-day exploits are unknown, it is hard to defend against such threats. As a result, there will be a continuous race between attacking and defending parties on discovering zero-day exploits.

SQL injection

SQL injection is an attack vector on web-based systems. In a SQL injection attack the attacker provide the target system with a malicious command or query. The hostile input data can trick the interpreting system into acting unintended resulting in privilege escalation or information disclosure. Injection attacks are at the top of the OWASP ”Top Ten Web Application Security Risks” list (OWASP, 2020).

3.2 Cybercrime-as-a-Service and Crime-Ware Markets

In parallel with the rise of Internet and legal e-commerce markets, e.g. Amazon, spawned dark web illicit markets. Among the most infamous dark web markets were Silk Road and Dream Market offering illicit drugs, arms, stolen credentials, child exploitation and cybercrime-as-a-service. The illicit markets mimic legal markets by providing vendor ranking, escrow services and quality assurances to enhance trust and efficiency according to Broadhurst et al. (2018).

The term *dark web* refers to a secret subset of the *deep web*. In contrast to websites on the regular *surface web* or *clearnet*, websites on the deep web is not indexed and as a result can not be found by the common search engines.

Early crime-ware markets described as: *”...a varied landscape of discrete, ad hoc networks of individuals initially motivated by little more than ego and notoriety”* (Ablon et al., 2014), have developed into highly organized markets with financially driven actors. The development and increased sophistication of these markets were propelled by the introduction of The Onion Router (TOR) and cryptocurrency as a method of payment. TOR was originally developed by the US Naval Research Laboratory to secure communication with agents stationed abroad. Both TOR and cryptocurrency provide anonymity and resilience towards law enforcement, thus lowering the risk of getting apprehended while trading illegal products. The resulting anonymous, easy to use markets with an enhanced resilience offering cybercrime-as-a-service, have made cyberattacks a tool not only available to actors with subtle programming skills (Meland and Sindre, 2019).

McGuire (2012) states that up to 80% of cybercrime could be the result of some form of organized activity and Europol (2017) noted a significant increase from 2013 in the number of criminal networks highly dependent on internet as part of their modus operandi or business.

Meland and Sindre (2019) argue that the popularity of malicious digital goods may indicate the type and capability of potential attackers, what assets they target and which vulnerabilities they are likely to exploit. Hence, it is highly interesting to review which crimeware products that are available and their popularity on dark web markets.

Dream Market was one of the largest known general crypto-market operating from 2013 to April 30th 2019 with 1800 unique vendors offering up to 100 000 illicit products for sale on a regular day. On April 14th 2018 the products on Dream Market were 51.1% illicit drugs and paraphernalia and 41.6% digital products such as malware, hacking tools and stolen credentials. Services and ‘other’ made up the remaining products (Broadhurst et al., 2018).

On average Broadhurst et al. (2018) observed nearly 12,000 unique digital products offered on Dream Market at any time. The majority of the products for sale were compromised accounts and credit cards. Hacking tools made up for 10.3% av the products. Such tools include toolkits such as Spyeye, phishing kits and hacking tutorials. Table 3.1 presents the distribution of product types and their average price.

Digital Product Type	% of Unique Listings	Average Price USD
Compromised Account	42.4	$32
Credit Cards	29.2	$31
Hacking Tools	10.3	$5
Documents (passports, ect)	6.7	$508
Vulnerabilities and exploits	0.94	$5
Keylogger	0.73	$3
Ransomware	0.73	$44
Botnet and DDoS	0.72	$20
Trojan and Virus	0.65	$19

Table 3.1: Distribution of digital products offered on Dream Market

** Original costs were in AUD and have been converted to USD using the exchange rate of 15.11.2019

Wehinger (2011) and Meland et al. (2020) suggest that fake items and scams flourish on the dark net. Based on this, Meland and Sindre (2019) argue that a more accurate distribution of the popularity of cybercrime products is derived from the number of successful sales. Figure 3.2 shows the number of successful sales per product category from the Apollon, Berlusconi, Empire and Grey markets. It can be observed that the most sold

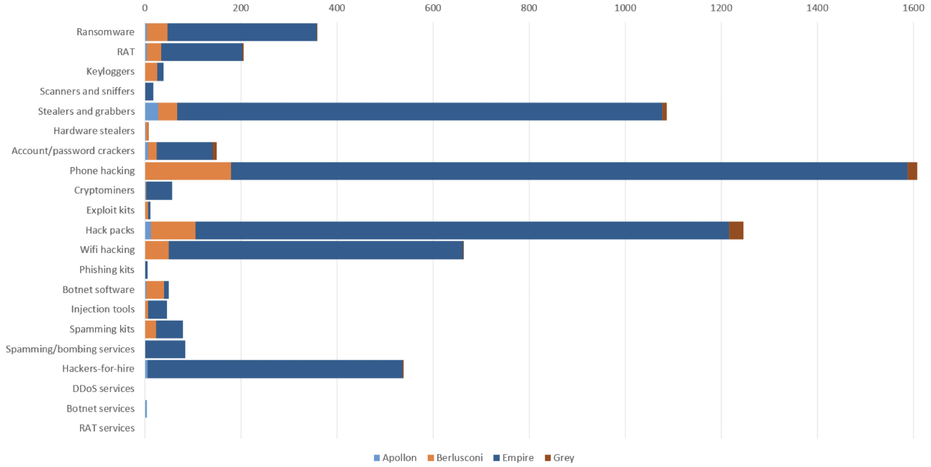


Figure 3.2: Figure collected from Meland and Sindre (2019) - Number of successful sales per digital product category. The categories are elaborated in Meland and Sindre (2019)

items are *phone hacking* (26%), *hack packs* (20%) and *stealers and grabbers* (17%). Here, *hack packs* refer to hacking tools and guides, while *stealers and grabbers* exploit clipboard data, e.g bitcoin addresses, usernames, passwords and bank accounts.

Further, Meland and Sindre (2019) present another way of looking at the marked by identifying the revenue of each product category. This is achieved by multiplying the number of successful sales with the latest listing price per item. The revenues are presented in figure 3.3 and show that *Hackers-for-hire* has the highest revenue due to a high item cost. Meland and Sindre (2019) note that the most sold items also tend to be the most expensive.

The study notes the general low cost and in particular the cost of ransomware tools. Knowing that the ransomware attack on Møller-Mærsk resulted in business losses in excess of USD $250-300 million according to Moller-Maersk (2017) and AGCS (2018), the USD $44 average cost of ransomware tools exemplifies the bias between attacker and victim expenses related to cyberattacks. Hence, there is a sustaining rational incentive of attack.

Finally, Broadhurst et al. (2018) have observed an increase in interactions between cybercriminals and state or quasi-state cybersecurity actors on hacker-markets. These interactions often involve offensive cyberoperations, and a result of the interaction is the increased value and sophistication of malware available to criminals, e.g. zero-day exploits.

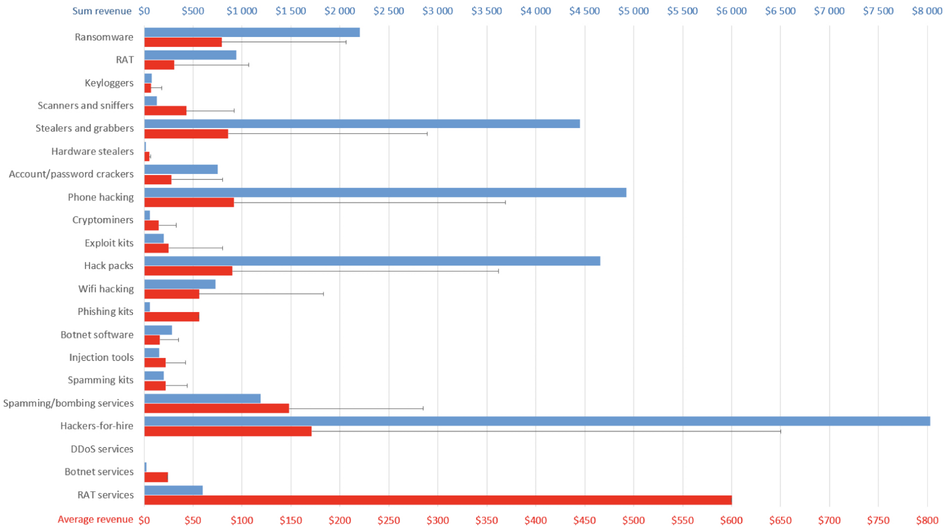


Figure 3.3: Figure from Meland and Sindre (2019) - Accumulated revenue per product category and average revenue per item from the Apollon, Berlusconi, Empire and Grey marketplaces

3.3 Criminal Behavior Models

A football team is considered ”good” or ”bad” depending on its ability to compete with other teams. This is analogous to a system or an organization ability to compete with the adversary on the cybercrime field. As the football manager studies the tactics and behavior of the opponent prior to a match, it is crucial to understand the cybercriminal in order to mitigate the cyberthreat.

3.3.1 Means, Motive, and Opportunity

It is well established in criminal investigation and a concept in criminal law that a person who committed a crime is likely to be a person who had a *motive*, *means* and *opportunity* (MMO) (Pendse, 2012), (Van Ruitenbeek et al., 2010). The MMO formulation states what encourages and facilitate criminal behavior.

Motive: Moorhead and Griffin (1998) defines a motive as a person’s reason for choosing one behavior from among several choices. In other words, why did the attacker launch a cyberattack?

Means: Pendse (2012) refers to means as the instrument(s) available to a person to carry out a task. The means are a source of power that gives the criminal the ability to commit the crime. In a cybercrime setting; which methods, tools and technology did the attacker use? An adversary possesses the means to attack when he or she possesses the capabilities to successfully execute the cyberattack.

Opportunity: McKendall and Wagner III (1997) define opportunity as the presence of a favorable combination of circumstances that makes an action possible. What made the criminal find the chance to commit the crime? Van Ruitenbeek et al. (2010) exemplify in a cybersecurity context that opportunity can be level of system access, system knowledge or attack skills needed to attempt the attack. Note that an adversary may have the ability to create an opportunity for attack, e.g. by gaining system access through a social engineering scheme.

3.3.2 Routine Activities Theory

In order to understand the adversaries of cyberassets, it is necessary to identify the rational incentives of attack. Cohoen and Felson (1979) present in their *Routine Activities Theory* a crime rationale analysis stating that crime will occur when there exist motivated offenders, suitable targets and the absence of capable guardians. Further, Cohoen and Felson (1979) argue that the lack of any one of these elements is sufficient to prevent a criminal act. Routine Activities Theory also claims that if the proportion of motivated offenders or suitable targets remain stable in a community, changes in routine activities, e.g. police patrol or network traffic analysis, could create more or reduce opportunities for crimes to occur.

Ekblom and Tiley (2000) and Grabosky (2001) extend Routine Activities Theory in a cybercrime context and state that cybercrime will occur when the following four conditions are met:

1. There exist an accessible and attractive target
2. The absence of a capable guardian
3. A motivated offender
4. The motivated offender possesses the required resources to commit the crime.

Here, the required resources is not just a question of technical skills, but also a requirement that the offender is able to invest in software development and hardware acquisition, as well as the time it takes to plan, prepare and perform the attack.

3.4 Cyberattack Modelling

3.4.1 The cyber kill chain

Already in 1998, Meadows (1998) presented a way of dividing attacks into different stages or phases to make visual representation easier. The next stage would not commence before the previous one had completed, and she used different colours to represent the assumed difficulty of each stage. The stages were not predetermined, but varied according to the nature of the attack. Later on, McQueen et al. (2006) defined a set of five fixed stages, *reconnaissance, breach, penetrate, escalation* and *damage*, which were then modelled as a compromise graph in order to find the weakest link(s) in the attack path based on expected time-to-compromise.

Similarly, Buldas et al. (2006) split attacks into two different phases, *preparation* and *break-in*, and model this in the form of an event tree.

Hutchins et al. (2011) describe different phase based models from military usage (countering terrorist attacks) and the information security field (between 2008-2010), and present their own version nicked the *intrusion kill chain*. This model was later on renamed and branded as the *cyber kill chain* (Hutchins et al., 2011) by Lockheed Martin, and has proven to be widely popular among defenders of IT and enterprise networks (Assante and Lee, 2015). The seven stages of the cyber kill chain are stated above in the Cyberattack section 3.1.1.

3.4.2 Attack tree cost modelling

Attack trees are acyclic graphs used to model threats from the viewpoint of the perpetrator. Schneier's original attack tree paper (Schneier, 1999) showed how different costs could be assigned to alternative leaf nodes and how these propagated to define the cheapest way of attack.

A fundamental paradigm for this kind of modelling is the assumption of a *rational attacker*. Buldas et al. (2006) define a rational attacker such that *1) there will be no attack if the attack is unprofitable* and *2) the attacker chooses the most profitable way of attacking*.

There have also been several approaches where costs are used in combination with other attributes. For instance, Buldas et al. (2006) include costs, gains, penalties and associated probability values. Further examples of different attributes and references to papers that utilize costs in attack trees is given by Bagnato et al. (2012). Having more attributes enables additional ways of analysing attack trees, for instance Kumar et al. (2015) show how to find the minimum time to complete an attack given a specific budget. Jensen et al. (2017) present an approach where cost is a function of time instead of a constant cost per atomic attack attempt.

Still, the major challenge of assigning accurate attribute values to attack tree nodes is difficult to overcome as attacker-specific information tends to be based on a best guess (Saini et al., 2008).

A comprehensive overview of more than thirty attack and defense modelling approaches based on directed acyclic graphs can be found in a survey paper by Kordy et al. (2014). A more recent survey focusing on fault and attack trees has been published by Nagaraju et al. (2017).

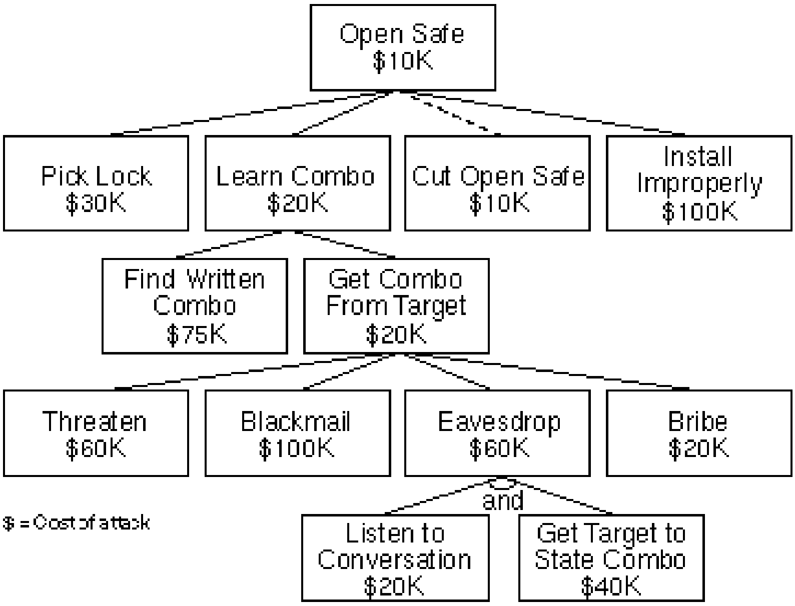


Figure 3.4: Attack Tree with cost attributes from Scheier's original paper on Attack Trees (Schneier, 1999)

Chapter 4

Literature Review - Cybercriminal Profiling

> ”If you know the enemy and know yourself, you need not fear the result of a hundred battles”

Sun Tzu - *The Art of War*

What was thought to be the key to winning a battle in the 6th century B.C. China, is still valid in today’s cybercrime world. In order to know the cyberenemy of today, the research field of cybersecurity has looked to traditional crime investigation techniques like criminal profiling. Criminal profiling aims to provide specific information regarding the type and characteristics of an individual who committed a particular crime (Jahankhani and Al-Nemrat, 2012). Shinder and Tittel (2002) state that a profile is a set of characteristics likely to be shared by criminals who commit a certain type of crime.

Kirwan and Power (2012) base the technique of criminal profiling to deduce behavioral characteristics on the two following assumptions:

- *The Consistency Assumption* - The premise that a criminal will exhibit similar behavior in all their crimes.
- *The Homology Assumption* - The premise that similar patterns of attack have to be associated with similar attacker characteristics and background.

Today, two methods of criminal profiling is prevalent: inductive and deductive profiling. *Inductive profiling* employs a database containing information on committed crimes and the characteristics of the offender who committed a particular crime. Utilizing the database, the profiler seeks to establish correlations between offender characteristics and specific types of crime (Shinder and Tittel, 2002). The method involves statistical analysis and pattern detection to identify matches between attacks and attackers.

Deductive profiling determines attacker characteristics based on analysis of forensic evidence collected from the crime scene and the principals of victimology - the study on the relation between the attacker and the target (Shinder and Tittel, 2002).

Truth to be told, criminal profiling is still based on educated guesses.

Inclusion criteria

Literature included in this review is on the topic of inductive attacker profiling. Inductive profiling is based on personal traits such as motivation, skills, ideology as well as the financial assets available to the attacker. These parameters can be associated with data on resources and resource alternatives in the IRCM, thus being relevant in the development of the attacker profiling feature of IRCM. For example, an attack requiring a high level of technical skills and capital indicates a nation state actor as a probable attacker. In contrast, an attack launched through running a script available on the Darkweb indicates a less sophisticated attacker.

Further, all the included literature is written in English and is either published by established publishers or grey literature published by internationally recognized organizations or enterprises.

It can be argued that the validity of research on criminal profiling do not expire only due to its year of publication. For this reason, we have not set a year for the literature to be published after in order to be included.

Exclusion criteria

Literature on deductive attacker profiling is excluded because forensic data is not available to IRCM. Forensic data, e.g. number of commands, speed of commands, programming errors and network measurements, is only available after an attack has occurred, while the IRCM is a tool to prevent an attack from occurring. Thus, forensic data and deductive profiling can not be utilized in the attacker profiling feature of the IRCM. Further, to limit the amount of literature to review, all literature with less than 10 citations were excluded.

To retrieve the literature we searched Google Scholar for the two terms: (1) ”inductive cyber criminal profiling” and (2) ”inductive cyber attacker profiling”. (1) gave 4,000 results and (2) gave 10,000 results both on 02.24.2020 when excluding citations. The review also used ”snowballing”, following up references in the reviewed literature, to identify relevant literature.

Due to the large number of results from the two Google Scholar searches, any literature without the terms ”cybercrime” or ”information security” in the headline was immediately excluded.

The literature reviewed in depth in this chapter is presented in Table 4.1 and supporting literature in presented in Table 4.2.

Title	Reference	Source
Proposed Methodology for Cyber Criminal Profiling	Warikoo (2014)	(1)
The Psyche of Cybercriminals: A Psycho-Social Perspective	Rogers (2011)	Referenced by Warikoo (2014) and Google Scholar search on title 03.16.2020 - 1 result
STIX^{TM} Version 2.0. Part 1: STIX Core Concepts	Jordan et al. (2017)	Handed out by co-supervisor PHM
Understanding Cyberthreat Motivations to Improve Defence	Casey (2015)	Referred to by Jordan et al. (2017) and Google Scholar search on title 17.09.2020 - 1 result
Threat Agent Library Helps Identify Information Security Risks	Casey (2007)	Referred to by Jordan et al. (2017) and Google Scholar search on title 17.09.2020 - 1 result

Table 4.1: Reviewed Literature

Title	Reference	Source	Reason of exclusion from in depth review
Examination of Cyber-criminal Behaviour	Jahankhani and Al-Nemrat (2012)	(1)	The paper does not propose any profiling method or criminal profiles. The paper summarize other proposed cyber criminal profiling methods
Towards a Methodology for Profiling Cyber Criminals	Kwan et al. (2008)	(1)	The paper proposes a deductive profiling method based on network analysis and other forensic data
The role of criminal profiling in the computer forensics process	Rogers (2003)	(1)	The article discusses the differences in inductive and deductive profiling in addition to the role that profiling can play in the computer forensic process. The article does not present any methodology or profile taxonomy.
The role of behavioral research and profiling in malicious cyber insider investigations	Shaw (2006)	(1)	The article proposes an extended profiling of insiders. This pure focus on insiders is the reason for exclusion.
A taxonomy and comparison of computer security incidents from the commercial and government sectors	Kjaerland (2006)	(2)	The paper presents a taxonomy for cyberattacks and not for attacker profiles.

Table 4.2: Supporting Literature

4.1 Warikoo (2014)

Title: *Proposed Methodology for Cyber Criminal Profiling*
Author: Arun Warikoo
Journal: *Information Security Journal: A Global Perspective, 23:172-178*
Year: 2014
Publisher: *Taylor & Francis*

Warikoo (2014) proposes a hybrid criminal profiling model with initial deductive processes before performing statistical analysis to identify common patterns and characteristics. Digital forensics data is utilized to provide clues about attacker sophistication, motivation, tools used and vulnerabilities exposed. These parameters are then used in an inductive manner to derive an attacker profile.

The methodology employs six *Profile Identification Metrics* to determine the adversary's modus operandi, psychology and characteristics. The metrics presented are: Attack Signature, Attack Method, Motivation Level, Capability Factor, Attack Severity and Demographics.

1. **Attack Signature** is given by the tools used for the attack. Such information is derived from analysis of the digital forensic evidence, thus being an deductive profiling method. Warikoo (2014) claims that digital forensic evidence provides information on the nature of the attack signature. Further, the paper exemplifies this by stating that a zero-day attack, exploiting unknown vulnerabilities, implies customized code developed for this particular attack. Such customized code points to an attacker with advanced programming skills. In contrast, an attack exploiting a known vulnerability which can utilize ready to use tools and scrips not developed by the attacker, suggest a wider range profiles with varying level of technical expertise.

2. **Attack Method** refers to the method used for the attack. Common attack methods include social engineering, distributed denial of service (DDoS), phishing and malware.

3. **Motivation Level** tells us how determined the adversary is to harm the victim. Warikoo (2014) states that the motivation level can be determined by the complexity of the attack. An attack with a high level of complexity where the offender has to exploit vulnerabilities on multiple layers and possibly combine several attack methods, e.g. social engineering followed by a malware based attack, indicates a persistent, highly motivated attacker. The paper also claims that such attackers are risk takers. Further, an attacker with a medium motivation level will not perform continuous attacks and a low complexity attack indicates a risk averse and nonpersistent attacker.

4. **Capability Factor** is defined by the availability of hacking tools, the ability to use those tools and techniques, as well as the resources at the disposal of the adversary. For instance, an unskilled or basic skilled attacker makes use of freely available tools. Further, an intermediate skilled offender additionally may purchase malware and has a handle on how the tools work and an advanced attacker is an expert in developing customized code for zero-day exploits.

5. **Attack Severity** states the impact an attack has on the victim. Warikoo (2014) classifies severity of an attack into the following:

   - Low: no tangible to the victim.
   - Medium: there is a moderate disruption on the victim
   - Major: major breach that can have major business impact
   - Critical: the victim goes out of business the moment the threat is realized.

6. **Demographics** and especially location is pointed out as a critical metric. The paper illustares this by claiming that cybercrime related to espionage have been known to originate from China.

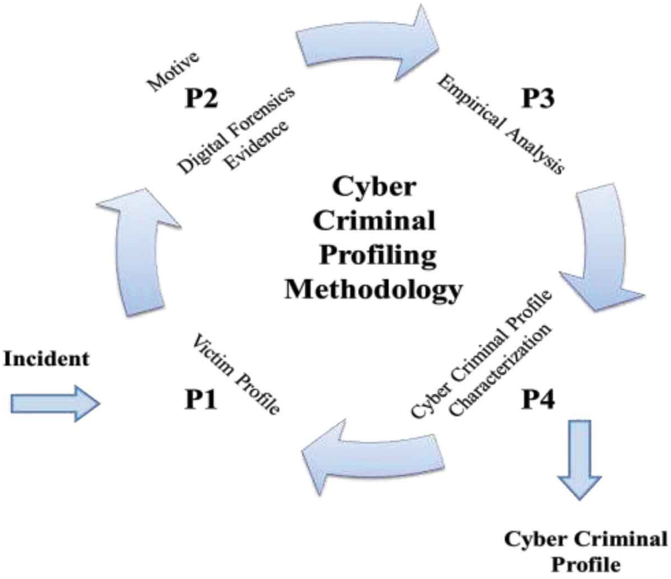

Motive
P2
Digital Forensics Evidence
P3
Empirical Analysis
Cyber Criminal Profiling Methodology
Incident
Victim Profile
P1
Cyber Criminal Profile Characterization
P4
Cyber Criminal Profile

Figure 4.1: Iterative Profiling Methodology proposed by Warikoo (2014)

The paper presents a four-step, iterative process, see Figure 4.1, to derive an attacker profile. The first step, P1, of the methodology involves victim profiling, e.i. identifying the aspects of the victim that attracted the attacker. Secondly, in P2 the profiling process involves identifying the motive behind the attack. Warikoo (2014) argues that the motive is closely associated with the victim. P2 relies on forensic evidence as well as inductive profiling. The third step, P3, involves empirical analysis on collected data and identifying characteristics by statistical analysis on the data against the metrics above. Finally, in the fourth step, P4, the results of the inductive profiling in the previous step are used to build a criminal profile which is classified into six profiles given in Table 4.3.

Cybercriminal Profile	Motive	Structure	Motivation Level	Skill Level	Attack Severity	Attack Method
Novice	Fun, Thrill	Unorganized	Low	Basic	Low to Medium	Freely available tools
Hacktivist	Political Activism	Unorganized	High	Basic to Intermediate	Low to Medium	Phishing, Spamming, DoS
Cybercriminal	Financial Gain	Unorganized with some level of collaboration	Medium	Intermediate	Medium to High	Spamming, Malware
Cyber Crime Syndicates	Financial Gain	Organized, Well funded	High	Intermediate to Advanced	High	Malware available on Darkweb markets
Cyber Spies	Espionage, IP Theft	Highly Organized, State sponsored	High	Highly Advanced	Critical	Customized code, Zero-day exploits
Cyber Terrorists	Disruption	Work in small modules, Well funded	Medium	Basic to Intermediate	Low to Medium	DoS

Table 4.3: Attacker Profiles proposed by Warikoo (2014)

4.2 Rogers (2011)

Title: *The psyche of cybercriminals: A psycho-social perspective*
Author: Marcus K Rogers
Book: *Cybercrimes: A multidisciplinary analysis*
Year: 2011
Publisher: *Springer*

Rogers (2011) seeks to answer the question: *”As supposedly rational beings, how is it that some of us choose to be involved in aberrant and destructive activities, including cybercrime?”*. The chapter presents a taxonomy of cyberoffenders to help develop profiles which identify common characteristics such as socio-demographics and personality traits that are likely to be correlated with different types of cybercrime. A taxonomy is a scientific classification by a pre-determined system. Here, the classification of an attacker is determined by technical expertise, overt behaviors, motivation and moral development.

Rogers (2011) presents seven categories describing archetype offender profiles. The seven categories are: script kiddies, cyber-punks, hacktivists, thieves, virus writers, professionals and cyber-terrorists. Figure 4.2 shows the categories placed in a continuum to reflect criminal behaviors that range from amateurs only interested in causing mischief, to state-sponsored terrorism. The model presents an increasing skill level left to right.

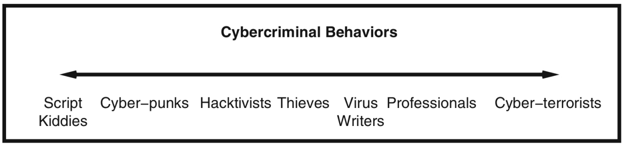


Figure 4.2: Rogers’ Taxonomy (Rogers, 2011) on cybercriminals. Attacker profiles are placed according to an increasing skill level left to right

Later, Rogers (2011) expands the one dimensional skill based model to include the motivation of attack in a circumplex model shown in Figure 4.3. This model includes four motivational quadrants: (1) Revenge, (2) Financial, (3) Notoriety, (4) Curiosity.

4.2.1 Rogers Taxonomy

In the presented taxonomy Rogers classifies sense of morality using *Lawrence Kohlberg’s stages of moral development*. Kohlberg (1974) states that our moral development has six proceeding, non-reversible stages. Each stage is more adequate at responding to moral dilemmas than its predecessor. Humans pass through these stages as we mature cognitively and morally. The norm for a well functioning adult is between stage 4 - *Authority and social-order maintaining orientation*, in short acceptance of authority - and stage 5 - *Social contract orientation* implying flexible judgements.

Script Kiddies (SK)

Rogers refers to *script kiddies* as individuals with technical skills limited to run precompiled software. Script kiddies seek to create mischief, without grasping the impact of executing the software. The primary motivators for this group are immaturity, ego boosting and thrill seeking. Script kiddies have an underdeveloped sense of morality with a self-interested orientation and seek attention and recognition from others. The SK brags about its exploits to others. This puts them at stage 2 according to Kohlberg (1974). The lack of sophistication and attention seeking make script kiddies the easiest group to apprehend by law enforcement.

Cyber-punks (CP)

Cyber-punks have a disrespect for authority and a disregard for societal norms. Their primary motivators are public attention and bragging rights associated with a successful attack. Thus, CPs are drawn towards cybercrime which generate maximum public attention, e.g. defacing webpages. The group do not fear getting arrested as this will generate publicity. A cyber-punk persona is a 12-18 year old male in Kohlberg's stage 2 that utilizes software developed by others.

Hacktivist (H)

The *Hacktivist* justifies its cybercrimes as civil disobedience and ascribing political and moral correctness. Rogers (2011) states that empirical data indicates that political motives usually is a secondary motive to revenge, power, greed, marketing or media attention. Hacktivists are placed between stage 2 - *Self-interest orientation* - and stage 3 - *Interpersonal accord and conformity*: following social norms - on Kohlberg's scale of moral.

Thieves (T) or Petty Criminals (PC)

This category include common criminals. They are motivated by money and greed, while avoiding attention as an effort to not being apprehended. The thieves and petty criminals target systems for financial gain. Common targets are credit card numbers, bank accounts and identity theft to be used in different types of fraudulent activities. These activities does not require sophisticated technical skills. On Kohlberg's scale of moral they are placed in stage 2 - *Self-interest oriented.*

Virus Writers (VW)

Virus Writers develop malware. Often, the individual that creates the virus, is not the one who releases it upon a target. Some virus writers may have a financial motivation and sell their malware on the Darkweb to buyers who categories as cyber-punks. Others in this category are driven by curiosity. As a result the VW can be placed from stage 2 to 5 on Kohlberg's scale of moral. Despite virus writers being a diverse group, they all posses above average technical skills.

Professionals (P)

For persons in the *Professional* category their cybercrime activities are their everyday job, thus they seek anonymity. They engage in competitive intelligence, e.g. corporate espionage, and sophisticated swindles. Morality nor ethics of their actions figure into the equation, the professional sells its services to the highest bidder. This places these individuals between the morality stage - 5 *Social contract orientation* - and 6 - *Universal ethical principles* where their ethics are dictated by self chosen principles.

Cyber-terrorist (CT)

Cyber-terrorists include military nation state actors and paramilitary agents. CTs are soldiers or freedom fighters on the cyberspace battlefield and their goal, like traditional military, is to win the battle by whatever means possible. A battle in cyberspace involves attacking or defending nation state assets, e.g. traditional military capabilities, infrastructure or the media in order to tilt elections and political decisions. The primary motivation of this group is to win the battle. Any moral stage is not relevant to this category.

4.2.2 Circumplex Model

Rogers (2011) expands the model in Figure 4.2 to include motivation of carrying out an attack, in addition to technical skills, as a parameter to explain attacker behavior. The four motivational quadrants are (1) Revenge, (2) Financial, (3) Notoriety and (4) Curiosity. In the circumplex model shown in Figure 4.3 the position of profiles relative to others represents the motivational component, while the position of a profile relative to the origin represent the skill level, e.i. the further from the origin, the more advanced skills.

In the expanded circumplex model the script kiddie term is coined *Novice* (NV), the hacktivists term is replaced by *Political activist* (PA), thieves are referred to as *Petty Criminals* (PT), professionals goes as *Professional Criminals* (PC) and the cyber-terrorist term is switched with *Information Warrior* (IW).

The new model also introduces two new categories: Internals (IN) and Old Guards (OG).

Internals (IN)

Internals or insiders are historically the group which represent the greatest risk and causes the most costly and destructive attacks. Primarily internals are disgruntled employees/ex-employees, consultants or contractors. The IN leverage their system access to cause harm on their own organization's systems. The fact that individuals in this group usually have elevated system access inherent to their position, e.g. IT-workers, in the organization imply above average technical skills. The primary motivation of this group is revenge upon their organization.

Old Guard (OG)

An *Old Guard* has no criminal intent, but is interested in the intellectual endeavor of cybercrime. OGs have advanced technical skills and, as virus writers, often don't leverage

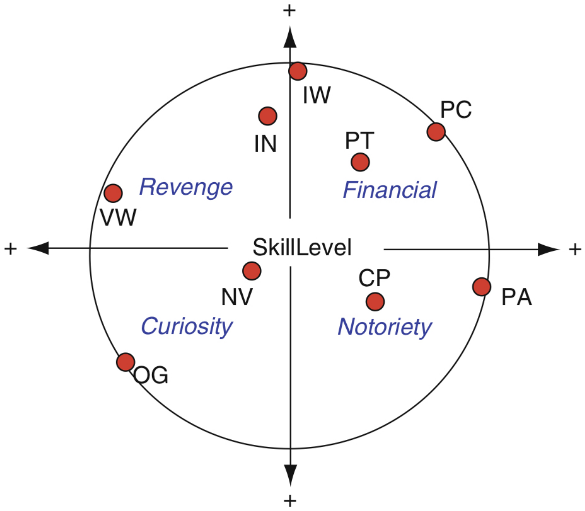


Figure 4.3: Rogers' (Rogers, 2011) circumplex model with profiles: novice (NV), cyber-punks (CP), petty thieves (PT), virus writers (VW), old guard hackers (OG), professional criminals (PC), information warriors (IW), and political activists (PA) are included as a discussion point only

the malware they create, but make them available to and encourage less skilled cybercriminals to use their malware. The primary motivation of OGs is curiosity and the intellectual challenge of creating the malware.

The taxonomy classifying the attacker profiles presented in Rogers (2011) is summarized in Table 4.4.

Cybercriminal Profile	Motives	Skill Level	Moral Stage	Targets	Methods
Novice (NV), Script Kiddie (SK)	Thrill seeking, create mischief, curiosity	Run precompiled software	2		
Cyber-punk (CP)	Notoriety	Utilize software developed by others	2	Well known web assets	Defacing webpages, DoS
Hacktivist (H), Political activist (PA)	Notoriety, revenge, greed, political views	Able to carry out attacks in addition to utilizing hackertools	2-3	Political organizations, controversial enterprises	Defacing webpages, DoS
Thieves (T), Petty criminals (PT)	Financial gain	Some technical knowledge, able to utilize software developed by others	2	Financial systems, credit card numbers, bank accounts, ID numbers	Wire transfer fraud, credit card fraud
Virus writers (VW)	Curiosity, Revenge, Financial	Able to create viruses and malware	2-5		
Professional (P), Professional Criminal (PC)	Financial	Advanced	5-6	Intellectual property, assets which can be turned into cash	All
Cyber-terrorist (CT), Information warrior (IW)	To win the cyberbattle	State of the art		Nation states, terrorist groups	Whatever means and methods necessary to win
Internals (IN)	Revenge	Above average		Organization of the attacker	Leverage elevated system access
Old Guards (OG)	Curiosity	Advanced			

Table 4.4: Rogers' taxonomy on cybercriminal profiles

4.3 Casey (2007), Casey (2015) and Jordan et al. (2017)

Title: *Threat agent library helps identify information security risks*
Author: Timothy Casey
Journal: *Intel White Paper*
Year: 2007
Publisher: *Intel*

Title: *Understanding cyber threat motivations to improve defense*
Author: Timothy Casey
Journal: *Intel White Paper*
Year: 2015
Publisher: *Intel*

Title: *STIX^{TM} Version 2.0. Part 1: STIX Core Concepts*
Editors: Bret Jordan, Rich Piazza and John Wounder
Manual: *STIX^{TM}*
Year: 2017
Publisher: *OASIS Commettee Specifications*

Casey (2007) presents a threat agent library of archetypal cybercriminal agents. The library is based on a taxonomy of eight attributes that uniquely define each malicious agent. The eight agent attributes are: *Intent*, *Access*, *Outcome*, *Limits*, *Resources*, *Skills*, *Objective* and *Visibility*.

The library presented in Casey (2007) is extended in Casey (2015) by adding a Motivation attribute identifying the driver that causes the offender to commit crimes. Casey (2015) argues that motivation indicates the nature of an expected attack, thus it has a significant impact on defense planning. As an example, by identifying the motives of an disgruntled employee such a threat actor can be intercepted before the situation becomes harmful.

Further, the motivation narrows probable targets to different agents. To illustrate, a mobster is likely to target assets that can easily be converted into cash, while an agent driven by notoriety will target visible assets bringing attention to the attacker.

Finally, Casey (2015) claims that motivation shapes the intensity and persistence of an attack. It is reasonable to assume that an agent driven by ideology will have the patience to launch a persistent attack, in contrast to a thrill and attention seeking agent launching a short-lived, intense attack.

Jordan et al. (2017) provide a vocabulary aggregating the taxonomies presented in Casey (2007) and Casey (2015). The aggregated taxonomy also adds a *Threat Actor Role* attribute and an extended, more thorough description of threat actor skill level and sophistication.

4.3.1 Attributes of the aggregated taxonomy

Intent

The *Intent* of the threat actor defines whether the agent intents to cause harm. The **Hostile** agent starts with the intent to harm and takes deliberate actions to achieve this. On the other hand, the **Non-Hostile** agent has the intent to protect its assets, but accidentally or mistakenly take actions that cause harm. A common such accident is to open a hostile spam email.

Access

Access defines the agents extent of access to the target's assets. An agent will either have **Internal** or **External** access to these assets.

Outcome

The primary goal of the threat agent is described in the *Outcome* attribute. The paper notes that methods used to accomplish the primary goal may have ancillary effects. Possible outcomes presented are:

- **Acquisition/Theft:** Stealing assets for resale or extortion.
- **Business Advantage:** Increased ability to compete by acquiring business processes or assets.
- **Damage:** Injure personnel, physical or digital assets, or intellectual property.
- **Embarrassment:** Public, unflattering portrayal of target. May cause the target to lose influence, credibility, competitiveness or stock value.
- **Technical Advantage:** Acquire production processes or production assets to improve a product of the attacker.

It is worth noting that the outcome of a non-hostile agent may be unintentional.

Limits

The *Limit* attribute describes the legal and ethical limits that may constrain an attacker. The presented options are:

- **Code of Conduct:** The agent follows applicable laws and codes of conduct associated with their field, e.g. a journalist not revealing it's source.
- **Legal:** Agents act within applicable laws.
- **Extra-legal, minor:** The threat actor may break laws in minor, non-violent ways, e.g. trespassing.
- **Extra-legal, major:** Agents with this legal limit have no constraints regarding breaking any laws. Members of organized crime typically take no such account of the law.

Skill Level

Skill Level describes the training and expertise an agent possesses.

- **None:** The agent have no expertise or training in the methods for attack, but still has the ability to carry out random acts of disruption or destruction.
- **Minimal:** The attacker can copy and utilize existing techniques and tools.
- **Operational:** These actors have an understanding of the underlying technology and methods used for an attack, in addition to being able to create new attacks. Jordan et al. (2017) use the term **intermediate** for this level of sophistication.
- **Adept:** The adept threat agent is an expert in technology and attack methods. An adept skill level implies the ability to apply existing attacks and create new attack vectors best suited for the particular attack. Jordan et al. (2017) use the term **advanced** for this level of sophistication.

Jordan et al. (2017) add three skill levels to their vocabulary of *Threat Actor Sophistication*, the equivalent of the skill level attribute in Casey (2007). The additional skill levels are:

- **Expert:** These threat actors are experts on security systems and focus on the discovery of and use of unknown malicious code. They are able to work with kernel mode rootkits and use datamining tools.
- **Innovator:** The innovator is typically a full time, professional criminal or state actor who is part of a well-funded organization. Innovators work in teams to discover zero-day vulnerabilities and exploits targeting any IT-system.
- **Strategic:** State actors who create vulnerabilities in commercial products with the intent of leveraging their vulnerabilities to influence product design, development and manufacturing, e.g create backdoors.

Resource

An agent *Resource* defines the organizational level of operation. This in turn determines the available resources. The resource attribute is linked to the skill level attribute through that the organizational level implies the skills at least available to the agent. The options for the resource attribute are:

- **Individual:** An independent agent with resources available to the average person. The minimum skill level is *none*.
- **Club:** Agents interact on a social and volunteer basis, e.g. agents exchanging tips on a blog. Agents within a club often have no personal interest in the specific target. At least a *minimum* skill level is implied.
- **Contest:** An interaction rooted in a common target. The interaction concludes when the goal is achieved and participants may be anonymous. A contest typically involves agents racing to break into a system for thrills and notoriety. Minimum skill level implied: *operational*.

- **Team:** A formally organized group with a leader. A team is organized around and motivated by a specific goal, in addition to operate within a single geography. A team implies *operational* skill level.

- **Organization:** Larger and more resourceful than a team. An organization typically operates in multiple geographies, persists long term and has an *adept* skill level at its disposal.

- **Government:** Controls public assets and functions within its jurisdiction, very well resourced which implies a minimum skill level of *adept*.

Objective

The *Objective* specifies the actions an agent intents to take in order to accomplish the desired Outcome. Casey (2007) gives the following options:

- **Copy:** Create a replica of the target asset such that the attacker gains simultaneous access to the asset.

- **Destroy:** Make the targeted asset worthless or unavailable to the victim.

- **Injure:** Damage an asset to an extent that limits its functionality or value, but the asset remains in possession of the victim.

- **Take:** The attacker take possession of the assets, simultaneous making the asset unavailable to the victim.

- **Don't Care:** When launching the attack the adversary does not have any rational plan, but it might make opportunistic choices during the attack.

Visibility

The *Visibility* attribute defines the extent to which the offender intends to conceal or reveal its identity. The presented options are:

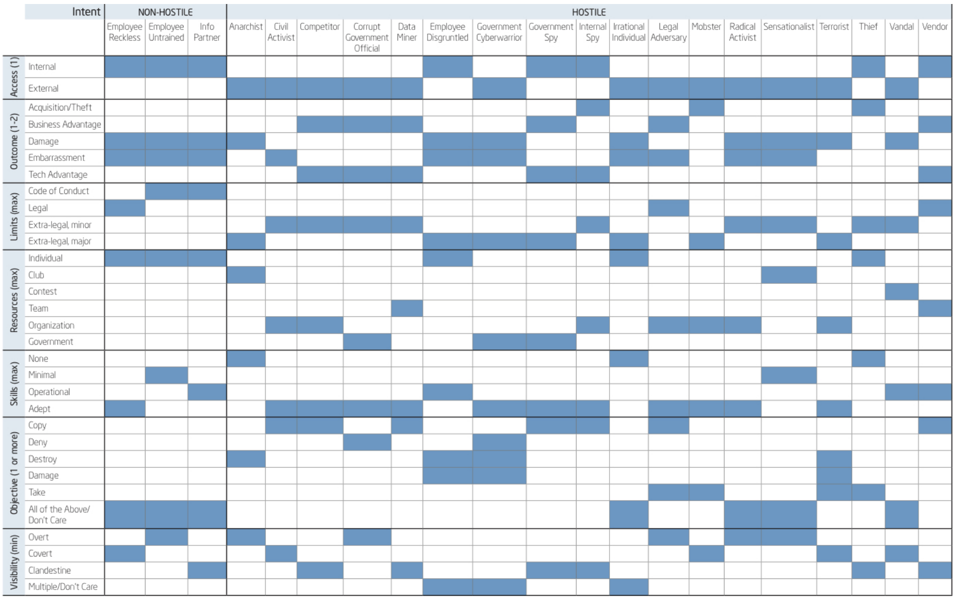
- **Overt:** Deliberately make the attack and the identity of the attacker known to the target before or at the time of attack.

- **Covert:** The agent intents to be unknown, but the target is aware of the attack at the time of execution or shortly after.

- **Clandestine:** The offender intents to keep both the attack and its identity secret.

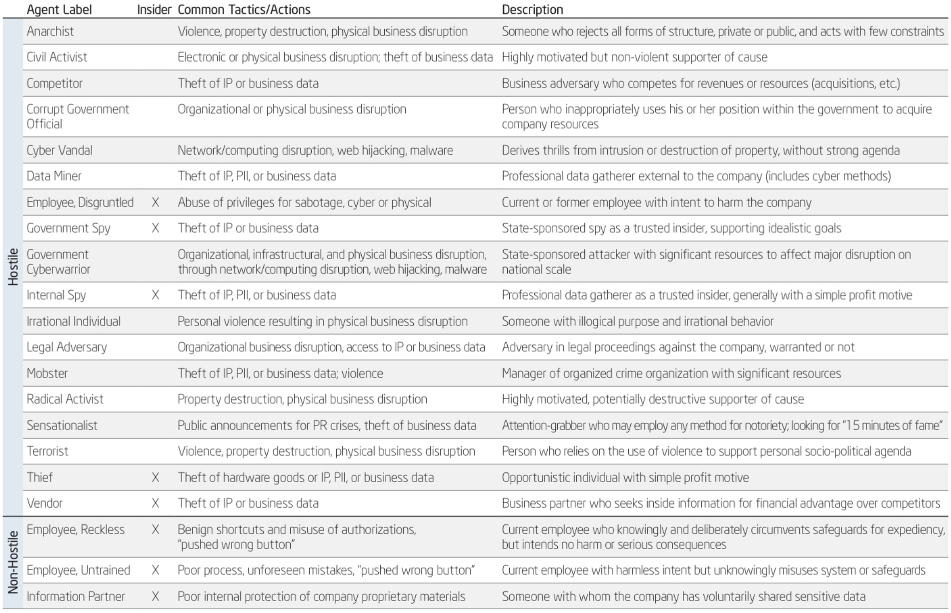
- **Don't Care:** The agent places no importance on secrecy or doesn't have a rational plan regarding visibility

The taxonomy presented in Casey (2007) is summarized in the tables shown in Figure 4.4 and Figure 4.5. Jordan et al. (2017) pinpoint that attacker profiles are not mutually exclusive. An adversary can both be a disgruntled employee and a spy.



	Intent	NON-HOSTILE			HOSTILE																	
		Employee Reckless	Employee Untrained	Info Partner	Anarchist	Civil Activist	Competitor	Corrupt Government Official	Data Miner	Employee Disgruntled	Government Cyberwarrior	Government Spy	Internal Spy	Irrational Individual	Legal Adversary	Mobster	Radical Activist	Sensationalist	Terrorist	Thief	Vandal	Vendor
Access (1)	Internal	■	■	■						■		■	■							■		■
	External				■	■	■	■	■		■			■	■	■	■	■	■		■	
Outcome (1-2)	Acquisition/Theft												■			■				■		
	Business Advantage						■	■	■			■			■							■
	Damage	■	■	■	■					■	■			■			■	■	■		■	
	Embarrassment	■	■	■		■				■	■			■	■		■	■				
	Tech Advantage						■	■	■			■	■									■
Limits (max)	Code of Conduct		■	■																		
	Legal	■													■							■
	Extra-legal, minor					■	■	■	■				■				■	■		■	■	
	Extra-legal, major				■					■	■	■		■		■			■			
Resources (max)	Individual	■	■	■						■				■						■		
	Club				■													■				
	Contest																				■	
	Team								■													■
	Organization					■	■						■		■	■	■		■			
	Government							■			■	■										
Skills (max)	None				■									■						■		
	Minimal		■															■				
	Operational			■						■											■	■
	Adept	■				■	■	■	■		■	■	■		■	■	■		■			
Objective (1 or more)	Copy					■	■		■			■	■		■							■
	Deny							■			■											
	Destroy				■					■	■								■			
	Damage									■	■								■			
	Take														■	■			■	■		
	All of the Above/ Don't Care	■	■	■										■			■	■			■	
Visibility (min)	Overt		■		■			■							■		■	■				
	Covert	■				■										■			■		■	
	Clandestine			■			■		■			■	■							■		■
	Multiple/Don't Care									■	■			■								

Figure 4.4: Taxonomy presented by Casey (2007)



	Agent Label	Insider	Common Tactics/Actions	Description
Hostile	Anarchist		Violence, property destruction, physical business disruption	Someone who rejects all forms of structure, private or public, and acts with few constraints
	Civil Activist		Electronic or physical business disruption; theft of business data	Highly motivated but non-violent supporter of cause
	Competitor		Theft of IP or business data	Business adversary who competes for revenues or resources (acquisitions, etc.)
	Corrupt Government Official		Organizational or physical business disruption	Person who inappropriately uses his or her position within the government to acquire company resources
	Cyber Vandal		Network/computing disruption, web hijacking, malware	Derives thrills from intrusion or destruction of property, without strong agenda
	Data Miner		Theft of IP, PII, or business data	Professional data gatherer external to the company (includes cyber methods)
	Employee, Disgruntled	X	Abuse of privileges for sabotage, cyber or physical	Current or former employee with intent to harm the company
	Government Spy	X	Theft of IP or business data	State-sponsored spy as a trusted insider, supporting idealistic goals
	Government Cyberwarrior		Organizational, infrastructural, and physical business disruption, through network/computing disruption, web hijacking, malware	State-sponsored attacker with significant resources to affect major disruption on national scale
	Internal Spy	X	Theft of IP, PII, or business data	Professional data gatherer as a trusted insider, generally with a simple profit motive
	Irrational Individual		Personal violence resulting in physical business disruption	Someone with illogical purpose and irrational behavior
	Legal Adversary		Organizational business disruption, access to IP or business data	Adversary in legal proceedings against the company, warranted or not
	Mobster		Theft of IP, PII, or business data; violence	Manager of organized crime organization with significant resources
	Radical Activist		Property destruction, physical business disruption	Highly motivated, potentially destructive supporter of cause
	Sensationalist		Public announcements for PR crises, theft of business data	Attention-grabber who may employ any method for notoriety; looking for "15 minutes of fame"
	Terrorist		Violence, property destruction, physical business disruption	Person who relies on the use of violence to support personal socio-political agenda
	Thief	X	Theft of hardware goods or IP, PII, or business data	Opportunistic individual with simple profit motive
	Vendor	X	Theft of IP or business data	Business partner who seeks inside information for financial advantage over competitors
Non-Hostile	Employee, Reckless	X	Benign shortcuts and misuse of authorizations, "pushed wrong button"	Current employee who knowingly and deliberately circumvents safeguards for expediency, but intends no harm or serious consequences
	Employee, Untrained	X	Poor process, unforeseen mistakes, "pushed wrong button"	Current employee with harmless intent but unknowingly misuses system or safeguards
	Information Partner	X	Poor internal protection of company proprietary materials	Someone with whom the company has voluntarily shared sensitive data

Figure 4.5: Taxonomy presented by Casey (2007)

Motivation

Casey (2015) gives the word *motivation* two meanings: *cause*, the reason a person commits an act, or *drive*, describing the level of interest or intensity a person acts on. The paper describes five motivational aspects: *Defining, Co-motivation, Subordinate, Binding* and *Personal Motivation*. The defining motivation is intrinsic to the agent and is assigned to all agents. It is the only aspect required to define an agent. The other aspects are optional *Motivational Modifiers* to elaborate the motivation of different agents. The motivational aspects are presented in Figure 4.6.

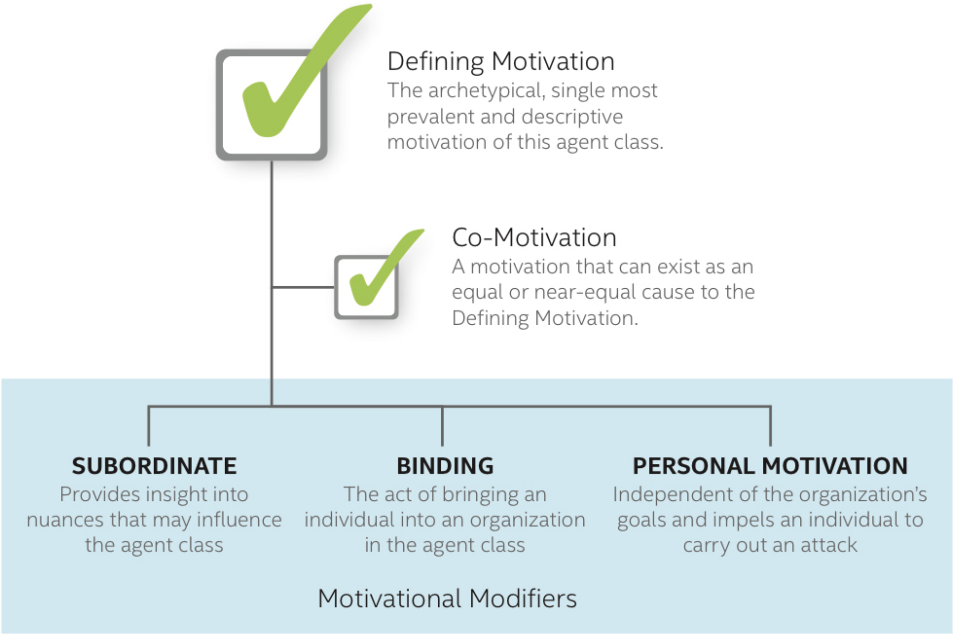

Defining Motivation
The archetypical, single most prevalent and descriptive motivation of this agent class.
Co-Motivation
A motivation that can exist as an equal or near-equal cause to the Defining Motivation.
SUBORDINATE
Provides insight into nuances that may influence the agent class
BINDING
The act of bringing an individual into an organization in the agent class
PERSONAL MOTIVATION
Independent of the organization's goals and impels an individual to carry out an attack
Motivational Modifiers

Figure 4.6: Motivational aspects presented by Casey (2015) to describe agents. The modifiers provide insights into nuances to the summed up motivation of a threat agent

In addition, Casey (2015) states that the organizational motivation defines the cause and the target, while the personal motivation strongly influences the chosen harmful actions of the attacker that the defending party must prepare for.

For the extended taxonomy Casey (2015) defines ten motivational elements:

- **Accidental:** The actor has benevolent or harmless intent, but with actions that inadvertently cause harm. This motivational element describes the non-hostile agent who unintentionally inflicts damage or disruption.
- **Coercion:** The agent is forced to act illegally on behalf of another, beneficial party through intimidation or blackmail by the beneficial party.
- **Disgruntlement:** The agent has a desire of revenge upon the target. Disgruntlement implies an interaction between the attacker and the victim.

- **Dominance:** The attacker seeks superiority over the target by bullying the target into submission.

- **Ideology:** An agent driven by ideology has a passion to express its ideas, beliefs and values through harmful actions.

- **Notoriety:** Acts by the threat agent is motivated by the agent personal desire for recognition and respect within a community.

- **Organizational Gain:** Agents of an organization engage in malicious activities seeking an advantage over a competing organization.

- **Personal Financial Gain:** The agent actions are driven by the opportunity of improved financial status.

- **Personal Satisfaction:** The harmful acts of the attacker fulfill its emotional self-interest like thrill seeking or curiosity.

- **Unpredictable:** These agents act without any identifiable reason or purpose and will create unpredictable, bizarre events with no logical explanation to the victim.

Figure 4.7 summarizes the motivational elements driving the different threat actor profiles.

Threat Actor Role

Jordan et al. (2017) extend the taxonomy from Casey (2007) and Casey (2015) with a *Threat Actor Role* attribute. The attribute describes the different roles that an attacker can play. For instance, the agent who developed a malware is not always the one releasing the malware upon a target. A single threat actor may possess several roles simultaneously. The different roles presented are:

- **Agent:** Threat actor who executes attacks either on behalf of themselves or at the direction of others.

- **Director:** An actor who selects objectives and goals. The director directs the malicious activities to meet the objectives and achieve the goals.

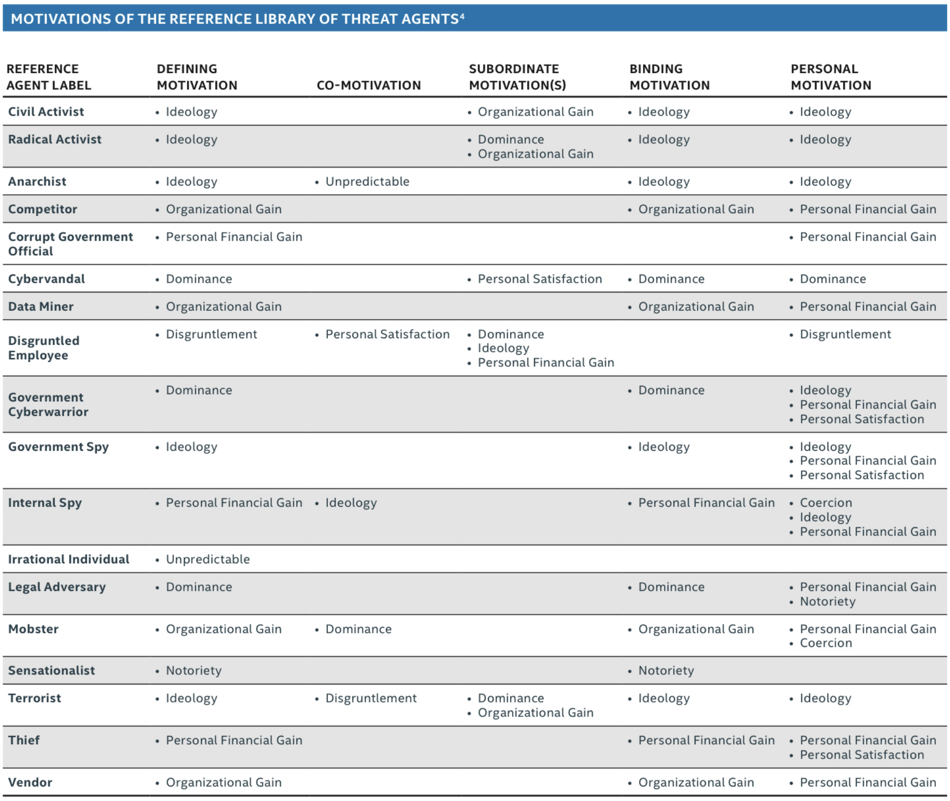
- **Independent:** Actors operating independently of others.

- **Infrastructure-architect:** An individual who designs the battle space.

- **Infrastructure-operator:** Threat actors who provide and support the infrastructure that deliver an attack, e.g. botnets and cloud services.

- **Malware-author:** These actors create malware, hackertools and discover unknown vulnerabilities.

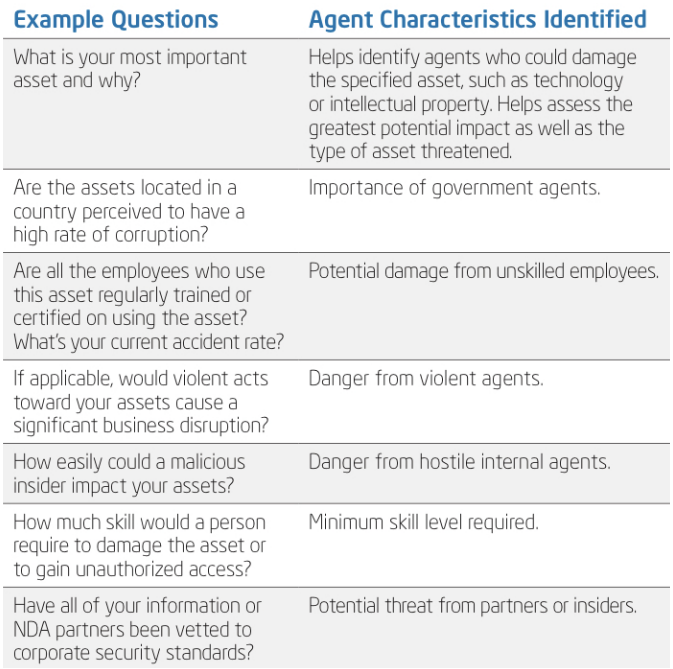
- **Sponsor:** Someone who funds malicious activity.



MOTIVATIONS OF THE REFERENCE LIBRARY OF THREAT AGENTS[4]

REFERENCE AGENT LABEL	DEFINING MOTIVATION	CO-MOTIVATION	SUBORDINATE MOTIVATION(S)	BINDING MOTIVATION	PERSONAL MOTIVATION
Civil Activist	• Ideology		• Organizational Gain	• Ideology	• Ideology
Radical Activist	• Ideology		• Dominance • Organizational Gain	• Ideology	• Ideology
Anarchist	• Ideology	• Unpredictable		• Ideology	• Ideology
Competitor	• Organizational Gain			• Organizational Gain	• Personal Financial Gain
Corrupt Government Official	• Personal Financial Gain				• Personal Financial Gain
Cybervandal	• Dominance		• Personal Satisfaction	• Dominance	• Dominance
Data Miner	• Organizational Gain			• Organizational Gain	• Personal Financial Gain
Disgruntled Employee	• Disgruntlement	• Personal Satisfaction	• Dominance • Ideology • Personal Financial Gain		• Disgruntlement
Government Cyberwarrior	• Dominance			• Dominance	• Ideology • Personal Financial Gain • Personal Satisfaction
Government Spy	• Ideology			• Ideology	• Ideology • Personal Financial Gain • Personal Satisfaction
Internal Spy	• Personal Financial Gain	• Ideology		• Personal Financial Gain	• Coercion • Ideology • Personal Financial Gain
Irrational Individual	• Unpredictable				
Legal Adversary	• Dominance			• Dominance	• Personal Financial Gain • Notoriety
Mobster	• Organizational Gain	• Dominance		• Organizational Gain	• Personal Financial Gain • Coercion
Sensationalist	• Notoriety			• Notoriety	
Terrorist	• Ideology	• Disgruntlement	• Dominance • Organizational Gain	• Ideology	• Ideology
Thief	• Personal Financial Gain			• Personal Financial Gain	• Personal Financial Gain • Personal Satisfaction
Vendor	• Organizational Gain			• Organizational Gain	• Personal Financial Gain

Figure 4.7: Motivational elements driving the different attacker profiles presented by Casey (2015)



Example Questions	Agent Characteristics Identified
What is your most important asset and why?	Helps identify agents who could damage the specified asset, such as technology or intellectual property. Helps assess the greatest potential impact as well as the type of asset threatened.
Are the assets located in a country perceived to have a high rate of corruption?	Importance of government agents.
Are all the employees who use this asset regularly trained or certified on using the asset? What's your current accident rate?	Potential damage from unskilled employees.
If applicable, would violent acts toward your assets cause a significant business disruption?	Danger from violent agents.
How easily could a malicious insider impact your assets?	Danger from hostile internal agents.
How much skill would a person require to damage the asset or to gain unauthorized access?	Minimum skill level required.
Have all of your information or NDA partners been vetted to corporate security standards?	Potential threat from partners or insiders.

Figure 4.8: Questions proposed to identify threat actor attributes in Casey (2007)

4.3.2 Identifying Threat Agents

Casey (2007) proposes an inductive methodology for identifying threat agents. Their method first identifies the attributes necessary to create a threat to an asset, before selecting threat agents form their library based on these attributes. For instance, the identification of the need for internal access to target the reviewed asset, implies all insider profiles such as disgruntled employee and internal spy. The paper pinpoints that the library consists of archetypes, thus an exact match between identified attributes and a profile is not possible or desirable. Figure 4.8 presents questions developed by Intel Threat Assessment Group to identify attributes of relevant agents. They suggest that the individual user of the taxonomy should develop their own set of question specific to them and their assets.

4.4 Summary

All the literature reviewed in depth present technical skills and motivation/motive as attributes that identify different threat actor profiles. In addition to agree on these two identifying attributes, the literature pinpoints that skills and motivation are the most descriptive attributes of the attacker profiles.

Warikoo (2014) alone explicitly states the relation between the attacker and the victim as an identifying attribute, while this relation falls under the motivation attribute in Rogers (2011) and Casey (2015).

Both Casey (2007), Casey (2015) and Warikoo (2014) suggest the level of motivation as a profiling parameter. Furthermore, the papers agree on structure/organizational level, in Casey (2007) referred to as ”Resource” of the attacking party, as an identifying attribute.

The method and type of attack is an attribute identified by both Warikoo (2014) and Rogers (2011). One can argue that the method and type of attack is to some extent covered by the skill level attribute presented in Casey (2007) and Jordan et al. (2017) through the fact that the different skill levels imply different attack methods.

Casey (2007), Jordan et al. (2017) and Rogers (2011) point to the type of target and the outcome and objective of the malicious actions as relevant, attacker descriptive attributes. They also agree on that ethics and moral separate threat agent profiles. This is suggested through the ”Limit” attribute of Casey (2007) and Rogers (2011) use of Kohlberg's scale of moral. Furthermore, these two groups of literature point out the attacker's desire of visibility as an identifying attribute.

Only Warikoo (2014) explicitly presents the technology and tools used for the attack, demography and impact of the attack as identifying attributes. It is worth noting that, to some extent the ”Objective” attribute in Casey (2007) overlap and describes the impact of an attack.

By contrast, Casey (2007), Casey (2015), Jordan et al. (2017) alone argue that the intent - hostile or non-hostile - and the attacker's level of access to assets - internal or external - are defining attributes of attacker profiles.

A comparison of presented identifying attacker profile attributes is summarized in Table 4.5.

Table 4.6 compares and summarizes the attacker profiles presented in the literature we reviewed in dept. Casey (2007) and Jordan et al. (2017) have a finer granulated set of attacker profiles, thus a larger number of profiles than the two other papers reviewed. It can be argued that many of the profiles presented by Warikoo (2014) and Rogers (2011) cover more than one of the more specific profiles in Casey (2007) and Jordan et al. (2017). In Table 4.6 each row shows profiles the study argues are comparable and in the case of a profile in Warikoo (2014) and/or Rogers (2011) covering a set of profiles in Casey (2007) and Jordan et al. (2017), the set of finer granulated profiles is put in the same table cell.

Identifying Attribute	Warikoo (2014)	Rogers (2011)	Casey (2007) Casey (2015) Jordan et al. (2017)
Method of attack	x	x	
Technology used	x		
Motivation	x	x	x
Skills	x	x	x
Impact of attack	x		
Demography	x		
Victim-Attack relation	x	x	x
Level of motivation	x		x
Structure / organizational level	x		x
Chosen target / outcome & objective		x	x
Moral stage / Limits		x	x
Intent			x
Access level			x
Visibility			x

Table 4.5: Summary of attacker profile identifying attributes

Warikoo (2014)	Rogers (2011)	Casey (2007) Casey (2015) Jordan et al. (2017)
Novice	Novice / Script Kiddie	
Hacktivist	Hacktivist	Civil activist Radical activist
Cyber criminal	Petty criminal	Thief
Cyber crime syndicate	Professional criminal	Data miner Internal Spy Mobster
Cyber spy	Information warrior	Government spy Government cyber warrior
Cyber terrorist		Terrorist
	Cyber-punk	Cyber vandal Sensationalist
	Virus writer	
	Internal	Employee disgruntled
	Old Guard	
		Employee reckless Employee untrained
		Anarchist
		Legal adversary
		Vendor
		Competitor
		Corrupt government official
		Irrational individual
		Information partner

Table 4.6: Summary and comparison of the attacker profiles presented in the reviewed literature

Chapter 5

Resource Cost Model

The *Resource Cost Model* (RCM) estimates the cost of the required resources to carry out a cyberattack. That is, what are the expenses of the attacking party measured in in dollars and cents related to completing an attack.

5.1 Cyber Kill Chain and Attack Trees

The *Intrusion Kill Chain* is a tool for describing each stage of a cyberattack. The kill chain analysis illustrates how an attacker must successfully and consecutively progress through all stages in order to complete a cyberattack. The seven stages of the kill chain are: *Reconnaissance*, *Weaponization*, *Delivery*, *Exploitation*, *Installation*, *Command and Control (C2)* and *Actions on Objectives (AoO)*. The objective of each stage is presented in Figure 5.1 and further elaborated in section 3.1.1.

In RCM each stage in the kill chain is coupled with a *Resource Tree* inspired by Schneier (1999) attack tree model. In RCM each *Resource Tree* derives the cost of the required resources at the given stage and have three levels: *kill chain stage*, *resource* and *resource alternative*.

The *resource level* defines which resources that are required to complete the kill chain stage stated in the root of the tree structure. Further, the *resource alternative level* presents the different alternatives available to the attacker in order to acquire the resource in the parent node. A resource alternative is a mean to realize its parent resource. Hence, to realize and acquire a resource, the attacker only needs to possess one of potentially multiple resource alternatives.

The essence of the model is that in order to carry out an attack, the adversary needs to complete all stages in the kill chain. Further, all resources required at each stage must be fulfilled in order to move to the next kill chain stage, while each resource only needs to be acquired through a single resource alternative. Hence, to mitigate an attack only a single resource must be made unavailable to the adversary by denying all resource alternatives.

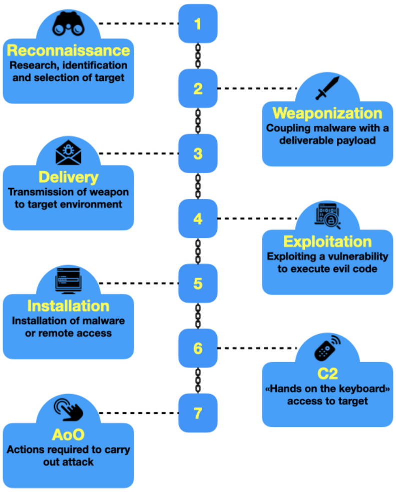

1
Reconnaissance
Research, identification and selection of target
2
Weaponization
Coupling malware with a deliverable payload
3
Delivery
Transmission of weapon to target environment
4
Exploitation
Exploiting a vulnerability to execute evil code
5
Installation
Installation of malware or remote access
6
C2
«Hands on the keyboard» access to target
7
AoO
Actions required to carry out attack

Figure 5.1: Cyber Kill Chain stages presented by Hutchins et al. (2011)

5.1.1 Resource Level

The *resource level* states which resources that are required to complete the parent kill chain stage. In RCM a resource is classified into five resource classes: *Skill*, *Tangible*, *Logic*, *Logic-atomic* or *Behavioral*.

Skill

Resources classified as a *Skill* are domain knowledge, the ability to develop a malware, follow guides explaining how to conduct known attacks or utilising ”working out of the box” cybercrime tools.

Tangible

Tangible resources include hardware components or other physical objects required. This can range from a laptop running Linux to a commercially available drone.

Logic

Commercially available software, data sets or cybercrime tools and cyberattacks available on the Darkweb are classified as a *Logic* resource. Examples of logic resources are ransomware kits and a Github repository.

Logic-atomic

Logic-atomic resources are resources that can not be broken into smaller parts without loss of meaning in the context of the attack, e.g an IP-address, email address or a password.

Behavioral

Behavioral resources include actions an agent must complete in order to carry out the attack, e.g. the victim opening a phishing mail or an attacker plugging in an USB-drive.

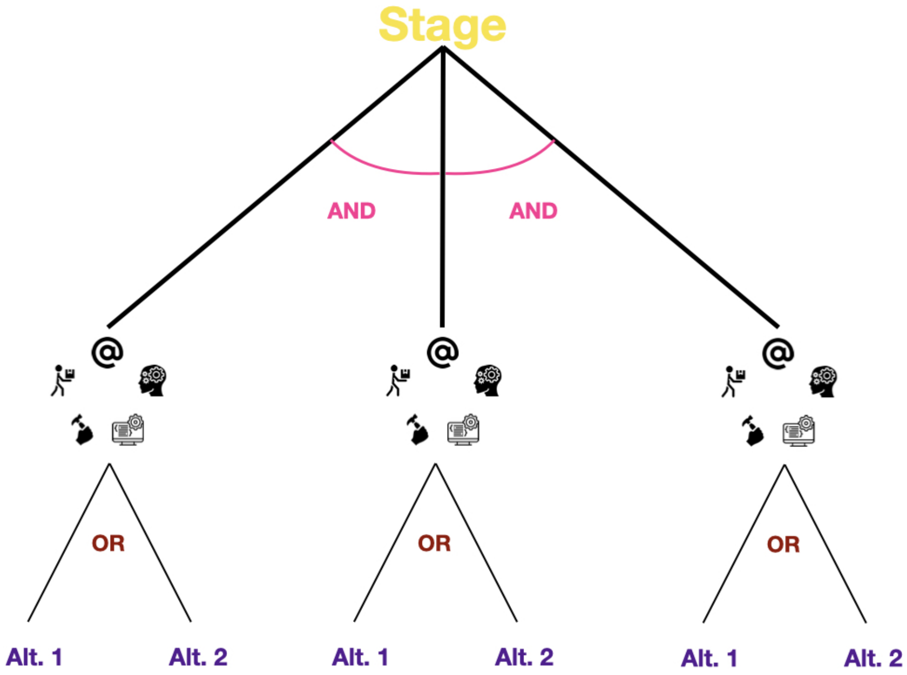


Figure 5.2: Resource tree - Each Kill Chain Stage is coupled with its own Resource Tree. The root node is the kill chain stage, the second level represents the resources and the leaf nodes are the resource alternatives.

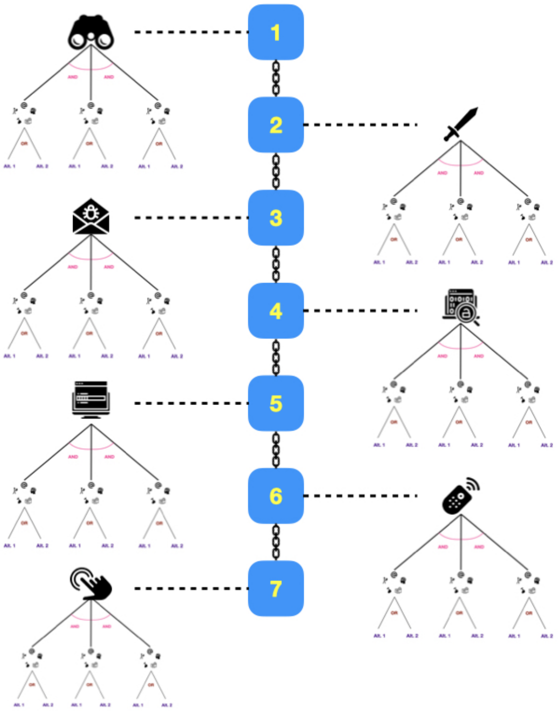


Figure 5.3: The seven kill chain stages coupled with resource trees

5.1.2 Resource Alternative Level

The *resource alternative* leaf nodes present a mean to realize its parent *resource* node. Each resource alternative is associated with a cost interval and a confidence. The cost interval states the estimated minimum and maximum cost of a resource alternative. For instance, a drone may be as cheap as $100 or as expensive as $1000. This result in the cost interval [$100, $1000].

The associated confidence represents the level of confidence the user assign to the cost interval and takes a value from 0 to 1. A confidence of 0 communicates that the user of the model has next to no evidence to support the stated cost interval and that the cost of the resource depend on many uncertain factors not known to the user. The cost estimate of bribing an insider will typically be associated with a low confidence, as the cost highly depends on the integrity and other traits of the insider.

In the other end of the scale, a confidence of 1 represents that the user has exhaustive evidence to back the stated cost interval and that the price of the resource alternative is not subject to great variation. Typically, any commercially over the shelf available hardware or software will be associated with a high confidence.

5.2 Estimating Cost

To launch an attack the adversary must complete all of the seven kill chain stages. Hence, the total cost of an attack is derived from the sum of the cost of the required resources at each stage. Further, the cost of a resource depends on the chosen resource alternative.

For an attack we say that we have a valid set V of resource alternatives when each resource alternative in V realizes a required resource such at all resources are realized. Let α represent the minimum estimated cost of the cheapest resource alternative and β represent maximum cost of the most expensive resource alternative of a resource. From this we derive the following estimated total cost interval:

$$total\ estimated\ cost = [minimum\ cost, maximum\ cost] \tag{5.1}$$

$$minimum\ cost = \sum_{\substack{\text{stage} \in \\ \text{kill chain}}} \sum_{i \in V} \alpha_i \tag{5.2}$$

$$maximum\ cost = \sum_{\substack{\text{stage} \in \\ \text{kill chain}}} \sum_{i \in V} \beta_i \tag{5.3}$$

By letting ϕ be the average confidence of the n resource alternatives associated with a resource R_j and c_i is the confidence of a resource alternative i associated with R_j, we get the following associated confidence C of the total estimated cost:

$$\phi_j = \frac{\sum_{i \in R_j} c_i}{n} \tag{5.4}$$

$$confidence = \prod_{\substack{\text{stage} \in \\ \text{kill chain}}} \prod_{R} \phi_j \tag{5.5}$$

5.3 Requirements for Minimum Viable Product implementation of RCM

From the description of the Resource Cost Model (RCM) above, the study derives the following requirements for a minimum viable product (MVP) of the Interactive Resource Cost Model (IRCM) web application:

1. Provide a satisfactory description of RCM such that users interpret the key concepts of the model enabling them to utilize IRCM to model a cyberattack.
2. Let the user couple each stage in the kill chain with a resource tree.
3. Enable the user to add an arbitrary number of resources to each stage.
    - Each resource should be associated with a resource class.
4. Enable the user to associate an arbitrary number of resource alternatives to each resource.

  - Each resource alternative should be assigned a minimum and maximum cost stating a cost interval which is assigned a confidence.

5. Functionality to edit and delete all added resources and resource alternatives.

6. Provide functionality estimating total cost and confidence.

In order to achieve a MVP all of the requirement above must be met, hence they are not prioritized in any manner.

Chapter 6

Cybercriminal Profiling in the Resource Cost Model

Jahankhani and Al-Nemrat (2012) define criminal profiling as a method to provide specific information on the type and characteristics of an individual who committed a particular crime. To elaborate, Shinder and Tittel (2002) state that a profile is a set of characteristics likely to be shared by criminals who commit a certain type of crime.

6.1 Resource Based Profiling

The *Resource Cost Model* presented in chapter 5 is developed to prepare, mitigate and raise awareness on the prevalent cyberrisk. This implies an inductive profiling methodology, as no forensic evidence to drive a deductive profiling scheme is present prior to an attack.

In the Literature Review in chapter 4, the literature agrees on *motivation*, *victim-attacker relation* and *technical skills* to be the most descriptive identifying attributes of a cybercriminal profile.

To accurately describe the motivation of an attacker or the victim-attacker relation through the required resources of a cyberattack is difficult. As an example it would be highly speculative to suggest that a certain set of resources imply revenge by an ex-employee as the primary motivation of an attacker. This difficulty of attribute description yields for many other profile identifying attributes presented by the literature, e.g. the intent of the attacker.

Further, the fact that there is no forensic evidence available to the *Resource Cost Model*, excludes profile attributes such as demography, technology used and method of attack to be accurately recognized. As a consequence, a resource driven, purely inductive profiling method will be inaccurate.

To meet these shortcomings of resource based profiling, while still providing value to the user of the *Resource Cost Model*, we present a set of resource based rules to exclude improbable attacker profiles from an initial pool of cybercriminal profiles. Rather than to derive the most probable cybercriminal profiles, our methods exclude profiles which are improbable based on the information available through the required resources presented to the *Resource Cost Model*.

6.1.1 Resource Based Identifying Attacker Profile Attributes

From the identifying attributes presented in the literature and the available information in the *Resource Cost Model*, the study's resource based cybercriminal profiling methodology utilizes the attributes: *Motivation*, *Technical Skills*, *Cost*, *Limits* and *Access*.

Motivation

The resource-based motivation attribute describes the time it will take an attacker to acquire or realize a resource alternative. As an example, the time it will take an attacker to develop a required hacking tool or to acquire an illegal GPS jammer from the black marked.

As mentioned, no set of resources will be able to accurately describe the motivation attribute extensively. However, it can be argued that the amount of time it will take to carry out an attack describes how important, i.e. level of motivation, the completion of the attack is to the cybercriminal. A time-consuming attack indicates a more motivated attacker than an attack which only require a minimal investment in time. In addition, a time-consuming attack may indicate that the target was deliberately chosen.

Technical Skills

The technical skill attribute describes the academic and technical level an attacker must possess to be able to realize the given resource alternative in the *Resource Cost Model*. To exemplify; in an attack that requires a ransomware the attacker can either buy a working off the shelf ransomware-kit or develop the ransomware from scratch. Arguably, the latter requires a more advanced skill set.

We differentiate the level of technical skills using the vocabulary presented by Casey (2007):

- **None:** The resource alternative requires no expertise or training to be realized.
- **Minimal:** The resource alternative can be realized through copying code and utilizing existing techniques and tools.
- **Operational:** The resource alternative requires an understanding of the underlying technology and methods used for an attack. The requirement to create new attacks falls into this category.
- **Adept:** The resource alternative requires an expertise in technology and attack methods to be realized. An adept skill level implies the attacker has the ability to apply existing attacks and create new attack vectors best suited for the particular attack.

Cost

In attacker profiling, the cost attribute is the same as the *Minimal Cost* of a resource alternative as described in chapter 5. The total minimal cost of an attack can eliminate cybercriminal profiles, e.g. an attack with a minimal cost of USD $1000 is unlikely to be launched by a script kiddie.

Limits

The limit attribute describes if a resource alternative can be realized legally. This attribute is deterministic - the resource alternative can either be realized **Legally** or **Illegally**.

An attack which requires an illegal resource alternative in order not to break the kill chain is argued to eliminate certain criminal profiles by Casey (2007) and Rogers (2011). As an example, a mobster already heavily involved in crime is not limited by an illegal resource alternative, in contrast to a non-hostile internal employee. One can argue that all hostile profiles are willing to act illegally as any cyberattack most likely is illegal. However, the study argues that there is a difference in the acceptance of the degree of illegal behavior between the act of launching a cyberattack using only legally acquired resources and to commit a crime in order to acquire an illegal resource prior to launching a cyberattack.

Access

The access attribute identifies if the resource alternative requires internal or external access level to the targeted system in order to be realized. The profiling method does not care how the attacker obtained internal access, just if it is necessary or not. Hence, the attribute can either take the value **Internal** or **External**.

6.2 Cybercriminal Profiles

In our initial pool of cybercriminal profiles, based on the profiles presented in the Literature Review in chapter 4, the study includes the following: *Script Kiddie*, *Hacktivist*, *Vandal*, *Petty Criminal*, *Mobster*, *Cyberwarrior*, *Terrorist*, *Internal - Hostile* and *Internal - Non-hostile*. Below each profile is associated with matching resource based identifying attribute values. For more in depth descriptions of the profiles see the Literature Review in chapter 4.

Script Kiddie (SK)

The *Script Kiddie* has a low level of motivation, thus time-consuming attacks are not attractive to this profile. The technical skills of the script kiddie are limited to *Minimal* and the profile only accepts a minimal cost. Script kiddies will only utilize resources that can be realized *Legally* and have *External* access.

Hacktivist (H)

Hacktivists have a medium to high level of motivation anchored in the political cause they represent, thus they may conduct time consuming, targeted attacks. The technical skills of a hacktivist are limited to *Minimal*. In order to fight for their cause, the hacktivist accepts some expenses. The hacktivist is willing to require resources *Illegally* and has *External* access level.

Vandal (V)

Vandals are low to medium motivated and will only invest a limited amount of time to conduct attention seeking attacks. The technical skills of the vandal are limited to *Minimal* and the profile accepts a low cost. Vandals will only utilize resources that can be realized *Legally* and have *External* access.

Petty Criminal (PC)

Petty Criminals have a medium motivation level, thus are willing to invest some time in attacks which bring financial gain. They possess *Operational* technical skills and accept a medium cost which is lower than the financial gain of a successful attack. The petty criminal is willing to require resources *Illegally* and has *External* access level.

Mobster (M)

Mobsters have a medium to high level of motivation, thus they may conduct time consuming attacks which bring financial gain. The technical skills at the disposal of the mobster are *Operational* and the profile accepts costly attacks as long as the possible financial gain exceeds the expenses. Mobsters won't second guess to require resources *Illegally* and have *External* access level.

Cyberwarrior (CW)

Cyberwarriors are state actors with a high motivation level, thus will conduct persistent, highly time consuming attacks. The cyberwarrior has *Adept* technical skills and he/she is able to launch any attack. In addition, the cyberwarrior is not limited by any cost and has at its disposal resources that may be required *Illegally*. As an immediate result of their *Adept* skill level, the cyberwarrior has *Internal* access.

Terrorist (T)

Terrorists are highly motivated and well-funded, thus they may conduct time consuming and costly cyberattacks to front their beliefs. The technical skills of terrorists are limited to *Minimal*. The terrorist is willing to require resources *Illegally* and has *External* access level.

Internal - Hostile (IN-H)

The *Internal-Hostile* profile has a medium motivation level and may launch attacks that require some time. The technical skills of the internal-hostile is *Operational*, as it is likely that such disgruntled persons who choose a cyberattack as their means of revenge have a technical background. The profile accepts some expenses, but is limited to *Legally* acquire resources. The fact that they are internals grant them *Internal* access level.

Internal - Non-hostile (IN-NH)

The *Internal-Non-hostile* profile describes internals which launch cyberattacks by accident, thus they are not motivated at all to invest any time or money in a cyberattack and will only possess resources that can be *Legally* realized. The fact that they are working in an environment which makes an accidental cyberattack possible yields an *Operational* skill level and an *Internal* access level.

6.3 Exclusion Rules

With the limited information on identifying attributes available from the *Resource Cost Model*, the study defines a rule set to exclude improbable profiles from our initial pool of cybercriminal profiles stated above.

There are two key points to the exclusion process:

1. If a resource is unavailable to an attacker profile, that profile is excluded form the set of probable cybercriminal profiles. Remember that if a resource in the *Resource Cost Model* is made unavailable, thus breaking the kill chain, the modelled attack can not be carried out. For a certain attacker profile this means that if a single resource is unavailable to that profile, either because of lack of motivation, limits in technical skills, cost, access or moral limitations, the kill chain is broken and the attack can not be carried out by that attacker profile. A resource is made unavailable to an attacker profile if it is forced chose a resource alternative with an associated attribute of a certain value. The result of a resource being unavailable to an attacker profile is that the profile is excluded.

2. If the accumulated value of the continuous attributes *Motivation* and *Cost* in a valid set of resource alternatives exceeds a threshold, one or more attacker profiles are excluded. As explained in chapter 5, a valid resource alternative set contains one resource alternative which realizes each resource required for the modelled cyberattack.

It is emphasised that only a single resource need to be unavailable to a profile for it to be excluded. As an example, an attack that requires only a single illegal resource will exclude a profile that is not willing to require resources *Illegally* purely on that single resource. Further, only one of the continuous attributes need to exceed the limit of a profile for the profile to be excluded.

Below the exclusion rules which each identifying attribute yields are defined.

6.3.1 Motivation Exclusion Rule

For the continuous *Motivation* attribute attacker profiles are excluded if the accumulated time to require all resources exceed a profile's motivational limit. More precise, a profile is excluded when the least time consuming set of valid resource alternatives exceed the motivation level of the profile.

Figure 6.1 shows the motivation limits of the different profiles on a continuous time axis. The points on the axis yield the following motivation:

- **No motivation** - no time invested in requiring resources.
- **Low motivation** - an effort of a working day, e.i. 8 hours.
- **Medium motivation** - an effort of a working week, e.i. 40 hours.
- **High motivation** - an effort of a year of work.

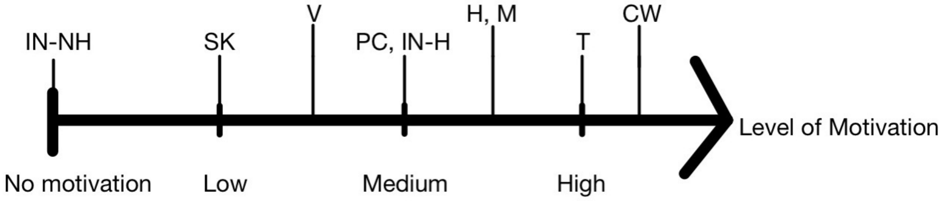


Figure 6.1: Limits of motivation of attacker profiles

	None	Minimal	Operational	Adept
Script Kiddie	X	X		
Hacktivist	X	X		
Vandal	X	X		
Petty Criminal	X	X	X	
Mobster	X	X	X	
Cyberwarrior	X	X	X	X
Terrorist	X	X		
Internal - Hostile	X	X	X	
Internal - Non-hostile	X	X	X	

Table 6.1: Technical Skill level of attacker profiles

6.3.2 Technical Skill Exclusion Rule

A profile is excluded due to lack of technical skills if any resource alternative required not to break the kill chain yields a higher *Technical Skill* level than the profile possesses. In other words, if the least technical skill demanding valid resource set requires a skill level more advanced than the profile possesses, the profile is excluded.

Table 6.1 shows which profiles are associated with the different discrete technical skill attribute values.

6.3.3 Cost Exclusion Rule

Profiles are excluded from the initial pool of all attacker profiles when the accumulated minimal cost of a valid resource set, i.e. *Total Minimal Cost* as elaborated in equation 5.2, exceeds the financial capacities associated with the profiles. We differentiate the financial capacities of attacker profiles on a continuous, logarithmic scale with the following thresholds:

- **No cost** - The total minimal cost of the attack is 0
- **Low cost** - The total minimal cost of the attack does not exceed USD \$100.
- **Medium cost** - The total minimal cost of the attack does not exceed USD \$1,000.
- **High cost** - The total minimal cost of the attack does not exceed USD \$10,000.

Figure 6.2 shows the financial capacities of the attacker profiles.

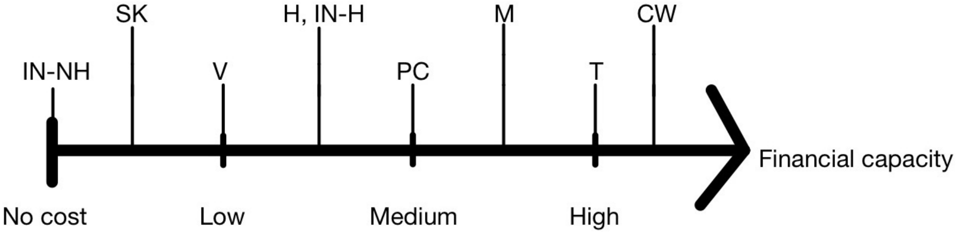


Figure 6.2: Financial capacities of attacker profiles

	Legally	**Illegally**
Script Kiddie	X	
Hacktivist		X
Vandal	X	
Petty Criminal		X
Mobster		X
Cyberwarrior		X
Terrorist		X
Internal - Hostile	X	
Internal - Non-hostile	X	

Table 6.2: Limits of attacker profiles to acquire or realize resources illegally

6.3.4 Limits Exclusion Rule

The *Limit* attribute is deterministic and its exclusion rule yields that if all valid resource alternative sets, i.e. resource alternative sets not breaking the kill chain, contain one or more resource alternatives which must be realized *Illegally*, all *Legally* limited profiles are excluded.

The moral limits of the attacker profiles are shown in Table 6.2.

6.3.5 Access Exclusion Rule

The *Access* level is deterministic and its exclusion rule yields that if all valid resource alternative sets contain one or more resource alternatives which require *Internal* access, all profiles with *External* access are excluded.

The access levels of the attacker profiles are shown in Table 6.3.

In context of the exclusion rules stated above the attacker profiles are summarized in Table 6.4.

	Internal	External
Script Kiddie		X
Hacktivist		X
Vandal		X
Petty Criminal		X
Mobster		X
Cyberwarrior	X	
Terrorist		X
Internal - Hostile	X	
Internal - Non-hostile	X	

Table 6.3: Associated Access level of attacker profiles

	Motivation	Technical Skills	Cost	Limit	Access
Script Kiddie	Low	Minimal	Minimal	Legally	External
Hacktivist	Medium-High	Minimal	Low-Medium	Illegally	External
Vandal	Low-Medium	Minimal	Low	Legally	External
Petty Criminal	Medium	Operational	Medium	Illegally	External
Mobster	Medium-High	Operational	Medium-High	Illegally	External
Cyberwarrior	High+	Adept	High+	Illegally	Internal
Terrorist	High	Minimal	High	Illegally	External
Internal - Hostile	Medium	Operational	Low-Medium	Legally	Internal
Internal - Non-hostile	No	Operational	No	Legally	Internal

Table 6.4: Summary of the identifying attribute values related to attacker profiles

Chapter 7

User Testing and Validation

The study conducted two rounds of user testing with three test groups. The user tests made up the final *Test* phase in the iterative design science workflow. User tests serve the purpose of validating solutions, identify issues and provide feedback from end users interacting with the artifacts being developed.

The first round of user tests tested a minimal viable product (MVP) of the Interactive Resource Cost Model (IRCM) with 8 scientists from SINTEF Digital with a cybersecurity background. The goal of the first round of testing was to validate that IRCM is able to convey the main concepts of the Resource Cost Model (RCM) and that users are able to construct RCMs. The first round of testing was conducted in February 2020 with in person observations of test subjects interacting with the model, followed by debrief interviews.

The second round of testing involved two separate test groups validating the usability of IRCM from the perspective of cybersecurity consultants and domain experts in maritime IT. Both groups tested the same version of IRCM. The main purpose of the tests with the cybersecurity consultants was to validate if IRCM provides any value when working with clients without cybersecurity experience, and the main goal of the expert tests was to validate that IRCM is able to model real world, full scale cyberattacks.

7.1 User Test 1 - UT1

In the first user test the study tested a minimum viable product (MVP) of the IRCM. Through early testing of an MVP the study can uncover if any of the basic features of the model is misinterpreted, confusing or misleading to the test subjects. It is critical to uncover such flaws at an early stage in order to correct them before more subtle tests. The later rounds of testing focused on what, if any, value the IRCM provides to end users rather than basic functionality.

The user tests of the MVP and the evaluation of these tests concluded the first loop in the iterative design science process.

UT1 was conducted in February 2020 with in person observations of interaction and debrief interviews regarding IRCM with 8 scientists from SINTEF Digital.

7.1.1 Blueprint for UT1

UT1 was conducted by providing the test subjects with a cyberattack case before asking the subjects to use the IRCM to derive a kill chain model. The case presented a GNSS cyberattack that is available in full in Appendix C.1. The test started at the IRCM web application ”Info”-page, see Figure D.1 in Appendix D, which provides information on how the model should be utilized and its features. This allowed us to test if the ”Info”-page had satisfactory information in order to derive a model of a known cyberattack. A test was successfully completed when the user had constructed a kill chain model with minimal guidance from the test executive. The study estimated in advance that a successful test would be completed within 15-20 minutes.

Situations where the subjects were in need of guidance identified issues with the current version and pinpointed specs/requirements for later versions.

The subjects were encouraged to think out loud during the test. This provided the study with valuable information on how they interpreted the IRCM basic functionality and how they utilized the web application to model a cyberattack.

After the tests were completed the test subjects took part in a debrief where they were encouraged to share their personal thoughts on weaknesses and strong suits of the IRCM. As the number of participants was not significant, it can be argued that a debrief generating qualitative data was preferable in opposition to a questionnaire generating quantitative data.

The whole session was estimated to take short of half an hour.

7.1.2 Test subjects

For UT1 the test subjects were volunteers from SINTEF Digital with a cybersecurity background and above average general digital skills. The study empathizes that none of the subjects had any training or interaction with the IRCM in advance. This was be the first time they encountered the model and its interactive implementation, e.i. IRCM.

In UT1 the study strove to test the intuitiveness of the application interface and its functionality. It can be argued that the cybersecurity background of the test subjects secured this. Issues raised during testing should as a result be due to weaknesses in the IRCM intuitiveness and interface, and not caused by the subject misunderstanding the case nor their lack of general computer skills.

7.1.3 Test environment

The tests were conducted on a general purpose laptop where the IRCM web application was running locally. Further, the tests were conducted in the offices of SINTEF Digital.

7.1.4 IRCM MVP

The IRCM MVP used during UT1 met all the requirements listed in 5.3 in addition to enabling the user to model an arbitrary number of cyberattacks. The modelled attacks are listed in the ”RC Model Index”-page shown in Figure D.2 in Appendix D. From this page the user has the option to view, edit or delete existing models in addition to create new Resource Cost Models.

The ”info”-page, see Figure D.1, strives to meet requirement 1. In addition to provide a short and precise description of RCM and an example of a resource tree, the page describes the different resource classes and how cost and confidence is estimated, before displaying a figure of the kill chain as a whole.

Although a bit technical and of less importance to some users, it can be argued that a formal description of how cost and confidence are calculated secures the trustworthiness of IRCM. The stated equations avoid users from perceiving the estimated cost and confidence as ”magic” numbers derived out of nowhere. Such magical numbers should be questioned by any woke user.

Each resource class is provided with an icon and a color. The use of icons and color coding is an instrument to enhance user’s memory of and convey the properties of the different resource classes. Hemenway (1982) shows that icons facilitate both initial comprehension and subsequent retention. Further, Wichmann et al. (2002) and Borges et al. (1977) show that the use of color enhance recognition. Icons and color coding are consistent on all pages in the IRCM in order to utilize and amplify their advantages. This follows Don Norman’s Design Principle of *Consistency* as defined by Norman (2002). Further, IRCM utilizes icons extensively to identify kill chain stages and navigation buttons for the reasons stated above.

Regarding the navigation icons, IRCM utilize standard icons such as a trash bin for delete actions. The usage of standard icons follows Don Norman’s Design Principle of *Affordance* (Norman, 2002), as the trash bin icon invites the user to perform a delete action. Users already associate a trash bin with delete actions from other computer applications, thus the icon serves as a clue to what the user can expect when clicking it. The same argumentation holds for the plus icon used for adding a resource or resource alternative, and the left arrows for returning to previously viewed pages.

Knight et al. (2009) show that icons might be interpreted differently across cultures and that users can associate unintended meaning and non-existent functionality to icons. The study recognizes these disadvantages and the use of icons is a subject to review and to validate through user testing.

The ”RC Model Show”-page, see Figure D.3, displays the RCM with all seven kill chain stages coupled with resource trees. The page also displays information provided by the user regarding the modelled attack and the current estimated cost estimate and its confidence. By clicking the kill chain icons, the user is brought to the ”Stage Show”-page, see Figure D.4, displaying the resource tree coupled with the clicked kill chain stage. This meet requirement 2. from section 5.3.

Further, the ”Stage Show”-page enables the user to add an arbitrary number of resources to each stage, meeting requirement 3. A resource is added by clicking the ”Add Resource” icon in the footer, which brings the user to the ”Add Resource”-page shown in Figure D.5.

From the ”Show Resource”-page, see Figure D.6, the user can view the details, edit and delete the displayed resource. This partly fulfills requirement 5. The footer on the ”Show Resource”-page offers the option to add a resource alternative to the displayed resource. Hence, checking of requirement 4.

By clicking the ”Add Alternative” icon in the footer, the ”Add Resource Alternative”-page, see Figure D.7 is shown. Here the user can describe the resource alternative, add estimated costs and confidence, thus providing the functionality described by requirement 4.

Further, by clicking on the name of a resource alternative either in the ”RC Model Show”-, ”Stage Show”- or ”Resource Show”-page the resource alternative and its associated values are shown by the ”Resource Alternative Show”-page, see Figure D.8. This page also provide functionality to delete or edit the viewed resource alternative, thus concluding requirement 5.

The study notes that when hoovering over any kill chain stage icon, resource or resource alternative name, the icon or text is highlighted and the immediate background changes color. This encourages the user to click the hovered item according to Norman (2002) principle of *Feedback*. This feedback is intended to convey the workflow of IRCM.

7.1.5 Observations

Reading through the ”Info-page”, which explains the concept of the Resource Cost Model, all test subjects stated a correct understanding of the main concepts of the model. They understood that to carry out the attack all stages must be completed and that the attack trees contain the resources required at each stage.

Five out of eight identified that in order to complete a stage all resources are required, while a resource needs only be realized through a single resource alternative.

After reading the ”Info-page” and the case described in Appendix C.1 was handed out, all participants were able to create a new model and populate stages with belonging resources and resource alternatives. To begin with when adding resources and alternatives, most test subjects jumped sporadically between the ”RC Model-show”, ”Resource-show”, ”Resource-add” and ”Alternative-add” pages, see Figures D.3, D.6, D.5 and D.7, in an exploratory manner, while saying that they were a bit confused regarding what is the intended work flow:

> ”What should I do now that I have added the first resource? Am I supposed to add an other resource or add alternatives?”

The intended workflow is to add a resource to a stage, then add all alternatives to that resource before adding another resource.

Throughout the tests the participants became more consistent and confident navigating the page. Seven picked up on the intended work flow.

One test subject tried to drag the ”Skill” icon on the ”RC Model-show” page, see Figure D.3, onto the ”Reconnaissance” icon in order to add a ”Skill” resource to the Reconnaissance stage. This presented an interesting potential workflow.

Although all test subjects were able to add resources and resource alternatives to the belonging stage, all but one struggled with distinguishing between a resource and a resource alternative. This was pinpointed by subjects misidentifying what are resources and what are its belonging alternatives. Six participants also expressed uncertainty when classifying resources and misclassified resources. Most misclassifications were rooted in the participant not remembering the definition of the different resources, then instead of navigating to the ”Info-page” to review the definitions they made a guess.

> ”I don’t remember exactly the definitions of the resource classes, but you can’t break a location into any smaller pieces that makes sense so I guess the location is a logic-atomic resource”

Subjects who revisited the ”Info-page” when adding resources were more secure and often, but not always, correct when choosing resource class.

Three out of eight test subjects expressed uncertainty that the confidence is a value in the interval (0,1), but remembered this when values outside this range were rejected and accompanied by an error message. One participant interpreted the confidence to describe how sure he was that the attacker could acquire the alternative, not the certainty of the estimated cost.

The creation of the models took more time than expected, up to 30 minutes.

7.1.6 Debrief interviews

All the test subjects stated that they see a potential value that the tool presents to actors trying to mitigate cyberthreats. Especially interesting is the fact that making only a single resource unavailable, thus breaking the kill chain, will mitigate the entire attack. This fact is believed to initiate and increase mitigation efforts as only making a single resource unavailable will be perceived as a doable and affordable effort. The perceived ease of mitigation is also thought to raise awareness.

Further, all the participants independently agreed that the structured visualization of the required resources will raise the awareness of the cyberthreat. The fact that many of these resources are fairly cheap and easy to acquire, is thought to illustrate how easily an attack can be launched. Some express that the fact that many resources are impossible to make unavailable will encourage the defending party to make the resources they control unavailable.

All of the participants expressed that the IRCM is a complex model which require that users take some time to fully understand it. A mentioned option is to provide new users with a live/video presentation explaining the model and how to use it, in addition to the existing information page. Further, all test users predicted that repeated use of the tool would increase their understanding of the model and confidence when using it. Six out of eight noted that they used IRCM with an increased confidence for every stage they completed. This is supported by the less sporadic page jumping observed.

The test subjects with prior knowledge of kill chains and attack trees expressed that this knowledge helped when trying to learn the concepts of IRCM.

Seven out of eight expressed that they didn’t fully understand the difference between a resource and a resource alternative. Further, they admitted that the amount of new information made them forget the definitions of the resource classes. Here, the subjects also pointed to repeated use as an advantage.

One participant commented that the current ”Info”-page is quite dense and packed with precise, scientific formulated information. The test user suggests to simplify the language at the expense of a longer information text.

An intriguing suggestion was to make the application into a game. It is thought that a gamification would make the users more creative when identifying resource alternatives, thus providing a more complete and correct cost estimate. The participant also suspected that a game will encourage more use as ”games are fun”.

Finally, it was suggested that for further testing the case presented to the users should be more in the lines of a story presenting how an attack is carried out on a concrete target, and not a general, formal presentation of an attack.

7.1.7 Results and Further Work

They main issue identified in UT1 was that the users didn’t recognize the difference between a resource and a resource alternative. To provide a satisfactory explanation of these two concepts and their relation is critical in order for users to be able to take use of IRCM.

A solution to this issue is to rewrite the explanation section on the ”Info”-page. As pointed out by a test subject, it might be productive to simplify the language and provide an extensive example. An option is to provide the users with a tutorial video explaining the model in depth.

The issues of users misclassifying resources and inserting invalid confidence values was rooted in users forgetting information provided on the ”Info”-page. This raises the requirement of repeated information on IRCM relevant to the different pages of the application. To be specific:

- Repeat information on resource classes at the ”Add Resource” page
- Repeat information on valid confidence values and what the different values represent at the ”Add Resource Alternative” page. A set of predefined values for the user to choose from could increase the user’s confidence in choice of value.

Finally, the interface should to a greater extent convey the intended workflow. This will have the intention to increase the confidence of navigation to novice users.

As suggested by a participant, information on where in the model/application hierarchy the user currently is, could be part of a solution. Further, an identified requirement is that the application should more actively suggest the intended workflow, while still providing the options to freely navigate the application. The ”Drag-and-Drop” workflow tried by a participant presents an alternative workflow.

Besides from identified issues and requirements regarding the IRCM, the study notes that a more concrete, story-like case should be used in further testing and that novice users take up to 30 minutes to create a complete model.

A gamification of the IRCM or to provide a gamified version in addition to the existing modelling tool, poses interesting further work in order to meet the overall requirement of raising the awareness of the persistent cyberthreat.

The above make up the following prioritized further work:

1. Clarify the distinction between a resource and a resource alternative.
   (a) Provide an example explaining the distinction between the to. The example must convey that a resource alternative is a mean to realize a resource.
   (b) Simplify the formulation of the description of the model on the ”Info”-page.

2. Repeat information on the IRCM relevant to the current page viewed by the users.
   (a) Repeat information on resource classes at the ”Add Resource” page
   (b) Repeat information on valid confidence values and what the different values represent at the ”Add Resource Alternative” page.
   (c) Add a set of predefined values for the user to choose from.
3. Convey intended workflow by actively suggest intended workflow.
4. Implement ”Drag-and-Drop” workflow.
5. Host application on remote server. Will allow multiple, simultaneous users to access application from their device.
6. Write a story-like case.
7. Implement automatic suggestion of potential attacker profiles of rational adversaries.
8. Make tutorial video explaining the model in depth.
9. Implement a gamification of the IRCM as an addition the current modelling tool.

7.2 User Test 2 - UT2

In the second round of user testing (UT2) in the design science workflow, the study tested the solutions for the issues identified during the first round of testing in addition to introduce the attacker profiling feature. With a validated model functionality form UT1, the main focus in UT2 was to validate if IRCM provides value to professional actors in the cyber domain.

UT2 had two separate test groups. Group A consisted of 8 cybersecurity consultants from Mnemonic AS. Group B was formed by 2 experts on maritime IT-systems from Navtor AS and SINTEF Ocean AS. All tests were conducted in May 2020 via video conference due to COVID-19 restrictions.

7.2.1 Blueprint for UT2

UT2 followed two different blueprints for testing of the two different test subject groups.

Blueprint UT2.A

UT2.A consisted of two phases; 1) test and validate the solutions to the issues identified in UT1 and 2) validate the value IRCM provides.

Phase 1 followed a similar blueprint as used during UT1. After asking the test subjects to read the ”Info”-page, see Figure D.10, the study asked the subjects to explain the difference between a resource and a resource alternative. This was intended to secure that the subjects had grasped the difference before beginning the modelling process. After the control question, the subjects were handed out a revised case describing the GNSS attack used in UT1, see Appendix C.2, before they were encouraged to build a Resource Cost Model. Phase 1 was concluded when the user independently and correctly identified resources and resource alternatives while constructing an RCM modelling the GNSS attack. A resource and its alternatives are correctly identified when the resource is assigned the correct resource class (Skill, Tangible, Logic, Logic-Atomic or Behavioral) and the alternatives are assigned reasonable values for cost, confidence, required skill and motivation, legal limit and access level.

The modelling process was cut off early in order to save time. When a user has understood the concepts and picked up on the workflow of the IRCM, further modelling is a repetitive process providing the study with minimal data. Phase 1 was estimated to take 30 minutes.

In order to observe the subject's interaction with IRCM, the subjects shared their screen in addition to the video link. Throughout the test session, subjects were encouraged to think out loud for the same reasons as in UT1. Further, they were only provided with guidance if getting completely stuck or repeatedly misinterpreting concepts or functionality of IRCM.

After phase 1 was completed, phase 2 consisted of a debrief evaluating their perception of IRCM and a discussion on what, if any, value IRCM provides.

To lead the discussion the subjects were asked the following questions:

1. What are your thoughts on modelling a cyberattack through the required resources?

2. How do you perceive the intuition of a resource-based modelling which is not talking in terms of technology?

3. Which, if any, value do IRCM provide?

4. How would you utilize IRCM in future work?

As the number of participants was similar to UT1, the study chose debriefing interviews over a questionnaire for the same reasons.

Blueprint UT2.B

In UT2.B the study consulted two domain experts in the maritime cybersecurity field in order to validate that IRCM is able to model a full scale, real world cyberattack. Further, the study wished to research if the domain experts see any value in using IRCM.

UT2.B consisted of three phases; 1) An interview where the expert presented the details of a cyberattack within the expert's field of expertise. During the interview the attack was reviewed in the context of the kill chain and the expert was asked to identify the required resources. Further, the expert was asked to elaborate on the cost of resource alternatives and values for the cybercriminal profile identifying attributes. The interview was estimated to take two hours. 2) After gathering information on the required resources and related resource alternatives, the study modelled the presented attack using IRCM. 3) With a complete Resource Cost Model (RCM), the study consulted the expert to validate the derived model and comment on the RCM and its accompanied implementation IRCM. The debrief was estimated to take one hour.

Although group A and B followed different testing blueprints, both are included in UT2, and not as two separate rounds of testing, as A and B tested the same version of IRCM.

7.2.2 Test subjects

The test subjects in UT2.A consisted of consultants from the Governance Risk and Compliance department of Mnemonic AS. All participants have multiple years of experience from the cybersecurity industry working with high profile clients. Their background and experience secure sound feedback regarding the value IRCM provides.

For UT2.B the study consulted two domain experts in the maritime cybersecurity field. One of the experts represents Navtor AS delivering solutions to digital navigation at sea and has an expertise on the maritime navigation system Electronic Chart Display and Information System (ECDIS). The Navtor employee validated the modelling of an attack on the ECDIS.

The other is a senior research scientist in the Transport and Energy department at SINTEF Ocean AS with expert knowledge on networks, their components and architecture, used on ships bridges. The SINTEF senior researcher validated the modelling of a cyber-attack on the administrative bridge network.

7.2.3 Environment

IRCM v1 was hosted on a Heroku web server. This allowed for remote user testing which deemed crucial due to COVID-19. Also, this met requirement 5 from UT1 further work, see section 7.1.7.

7.2.4 IRCM Version 1 Description

For an in depth description of the IRCM Version 1 (IRCM v1) development and deployment environment see Appendix E. In IRCM v1 the study introduces an attacker profiling feature based on the profiling methodology described in Chapter 6. The feature meets requirement 7 stated in 7.1.7. From the end user's perspective this mainly involves an extended ”Add Resource Alternative”-page, see Figure D.15, and the new ”Attacker Profile Info”-page shown in Figure D.9. The cybercriminal profile exclusion rules defined in 6.3 excluding improbable attacker profiles are utilized in the backend of IRCM v1 and their implementation is not visible to the end user.

In the extended ”Add Resource Alternative”-page the user can provide IRCM v1 with information on the identifying attributes *Motivation*, *Technical Skills*, *Limit* and *Access*. The identifying attributes used to exclude improbable cybercriminal profiles are elaborated in 6.1.

The new ”Attacker Profile Info”-page provides the user with information on the different cybercriminal profiles. It is intended as a reference for further information on attacker profiles that users consult after IRCM v1 has derived a set probable attacker profiles.

To conclude the user experience of the implemented cybercriminal profiling methodology, the ”RC Model Show”-page, see Figure D.11 and D.12, now displays the set of probable attacker profiles as an extension of the section with information regarding the modelled attack. The study chose to only display probable profiles in order to keep focus on these. Another option discussed was to display the whole set of attacker profiles and ”strike through” profiles as they were excluded. It can be argued that by only displaying the probable profiles the study secure that these profiles maintain in focus. Further by not displaying excluded, improbable profiles, these do not have the ability to grab the user's attention. This follows the saying: ”Out of sight, out of mind” and secures that the user prepares for attacks by the probable attacker profiles.

Further, IRCM v1 implements solutions to issues identified during UT1. Below it is described how the study solved these issues and met the design requirements stated in 7.1.7.

Clarify the distinction between a resource and its resource alternative

In order to meet the main identified requirement from UT1; to clarify the distinction between a resource and a resource alternative, IRCM v1 provides an example which explains the distinction in a non-cybersecurity context. The example, as part of the ”Info”-page, is shown in Figure 7.1. The updated ”Info”-page is shown in its entirety in Figure D.10. The study chose the non-cybersecurity ”Serve Cake” example because IRCM is a tool for users with and without a cybersecurity background. Hence, the cybersecurity aspect of an example in a cyber-context might steal the attention or confuse the non-cybersecurity schooled user.

Further, it can be argued that by pulling the user out of a cybersecurity mindset, IRCM eases the interpretation and amplify the distinction between a resource and a resource alternative. With the single focus on the distinction through a trivial example, the study secures that any user grasps this key concept. Any rational actor will agree that a birthday cake requires both a cake **and** lights, while the rational actor acquires the cake resource either by buying a cake **or** bake it itself.

To conclude the efforts to clarify the distinction between a resource and a resource alternative, the study simplified the formulations describing the model. This met the requirement 1.b) from the prioritized list in section 7.1.7. A concrete example is the switch from the term ”attack tree” to ”resource tree”. The ”attack tree” term was used as the tree structures in the model is inspired by Schneier (1999) attack trees. UT1 reviled that the word ”attack” mislead users from understanding that the tree structures model resources of a cyberattack and not the actions to conduct the attack.

Repeated information

Requirement 2 from the prioritized list of further work from UT1, see 7.1.7, states the need to repeat information from the ”Info”-page relevant to the current page viewed by the user. Especially, users had forgot the descriptions of the different resource classes when they added a resource. This resulted in resources being misclassified. To avoid misclassification IRCM v1 repeats the description of the resource classes on the ”Add Resource”-page. The change of the ”Add Resource”-page is shown in Figure 7.2.

Further, information regarding the meaning of the different confidence values is repeated on the ”Add Resource Alternative”-page, see Figure D.15. To secure correct input values and boost the confidence of the user when setting the confidence value, IRCM v1

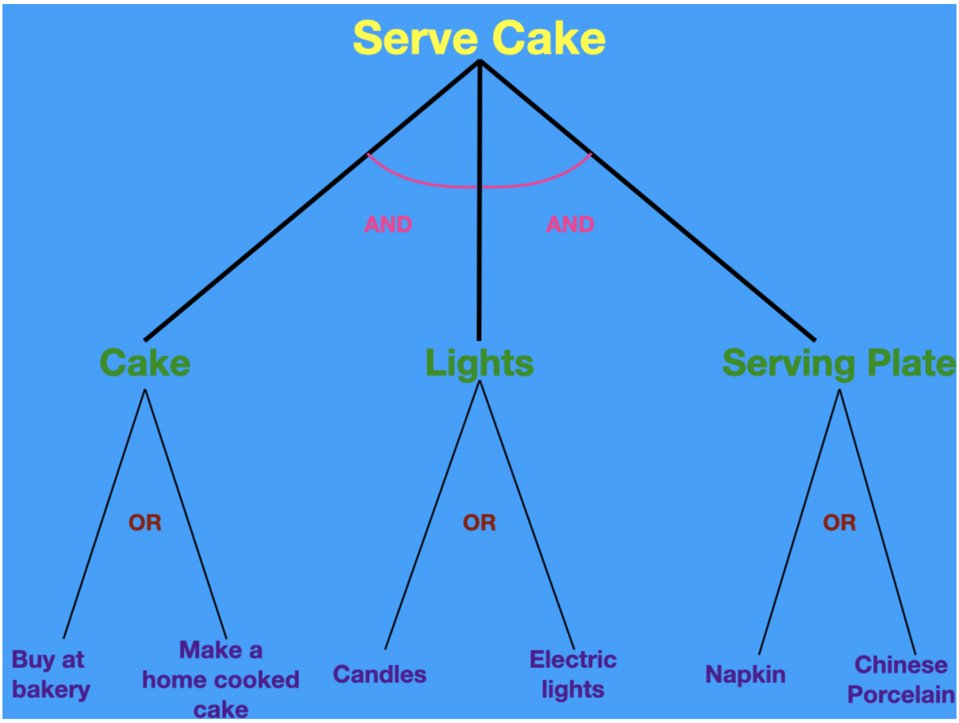
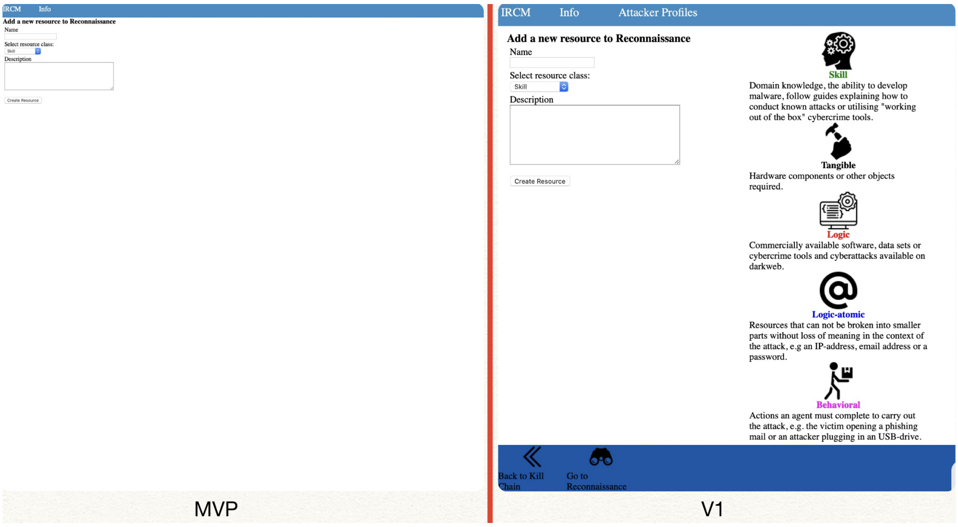


Figure 7.1: The ”Birthday party attack” - Every successful birthday party has a ”Serve cake” stage. The above Resource Tree defines the required resources of the stage and the resource alternatives.



Add a new resource to Reconnaissance

Name

Select resource class:

Skill

Description

Add a new resource to Reconnaissance

Name

Select resource class:

Skill

Description

Skill
Domain knowledge, the ability to develop malware, follow guides explaining how to conduct known attacks or utilising "working out of the box" cybercrime tools.

Tangible
Hardware components or other objects required.

Logic
Commercially available software, data sets or cybercrime tools and cyberattacks available on darkweb.

Logic-atomic
Resources that can not be broken into smaller parts without loss of meaning in the context of the attack, e.g an IP-address, email address or a password.

Behavioral
Actions an agent must complete to carry out the attack, e.g. the victim opening a phishing mail or an attacker plugging in an USB-drive.

Figure 7.2: The ”Add Resource” page in the MVP compared to V1

gives the user a set of predefined values to choose from rather than to type in the confidence value as in the MVP version.

Convey intended workflow

The study observed a sporadic, exploratory workflow among the users in UT1 and further work suggested to a greater extent convey the intended workflow: add a resource to a stage, then add all alternatives to that resource before adding another resource.

After ideating improved ways to convey the intended workflow, such as the ”drag-n-drop” suggested by a test subject in UT1, the study concluded that IRCM v1 should not force upon the user our intended workflow. Rather, IRCM v1 provides the user with more navigation options and leaves it to the user to find a preferred workflow. The added navigation options are shown in the footer in Figure D.14 and D.16.

It can be argued that by providing the user with extensive navigation options facilitating several workflows, IRCM v1 satisfy all the different workflows users might prefer. The fact that the final cost estimate and set of probable attackers are identical regardless of different workflows yields that there is no ”correct” way of modelling. An example of two different workflows is 1) first add all resources to all stages, before adding all resource alternatives and 2) add a resource to a stage, then add all alternatives to that resource before adding an other resource.

7.2.5 Observations

The most significant observation during UT2.A was that all 8 participants were able to correctly identify and classify resources and their belonging alternatives in the presented case. Meaning, they assigned the correct resource class to resources and assigned reasonable values to the resource alternative attributes. This was achieved with minimal guidance.

As observed in UT1, the participants in UT2 used some time to model the first resource and its resource alternative, but the speed of modelling increased as more resources and resource alternatives were added to the kill chain.

In advance of the second round of testing the study was unsure regarding the naming of the *Motivation* identifying attribute. In our opinion the term motivation is broader than just the amount of time an attacker is willing to invest. Based on this the study predicted the term might become subject to confusion among the users. The study discussed other names, such as ”Opportunity” in context of the MMO model described in 3.3.1, but in the lack of a better fit the study stayed with *Motivation*.

The observations from UT2.A show that all but one interpreted the motivation term and did not second guess it. The test subject that commented on the motivation term raised the concern that motivation describes more than just an investment in time. She interpreted motivation along the lines of revenge or greed, but she independently recognized that such feelings are difficult to deduce from the required resources of an attack.

3 out of 8 expressed uncertainty regarding the *Access* attribute. They did not interpret the attribute to describe the required level of access to the target system. Rather, they interpreted the access term to describe the access level of the specific resource alternative.

> ”You need to be enrolled at NTNU to participate in the signal processing course, thus the course requires internal access”

7.2.6 Debrief Interviews

UT2.A

Answering question 1 and 2 regarding the resource based model approach, three out of eight express that the Resource Cost Model is more tangible, thus more concrete on possible mitigation efforts than models with an action based approach.

> ”The approach of identifying what you need to possess, rather than what you need to do is more tangible. It seems more doable and concrete to make a required resource unavailable, than to deny an action in order to mitigate an attack.”

One subject states that the resources derived by IRCM provides an intuitive cost-benefit context for mitigation efforts.

> ”By determining the resources in a kill chain approach, IRCM makes it easy to evaluate the cost of making resources unavailable. This cost can then be compared to both the total cost of conducting an attack and its consequences”

Further, three participants independently agree that the cost interval is a solid parameter for expressing the probability of attack. Two of the three also anticipate that wealthy mobsters and well-funded state actors will try a cheap attack before moving to more sophisticated, expensive attacks. All three elaborate that with the cost interval expressing the probability of an attack, it becomes trivial to compare attacks and as a result focuses mitigation efforts on the most probable attacks.

> ”Regardless of the attacker, a cheap attack will always be more probable than an expensive one. By deriving the cost of launching an attack, IRCM quantifies a usually diffuse measurement of risk expressing probability”

Addressing the third question; ”Which, if any, value do IRCM provide?”, the test subjects emphasize the profiling feature and the structured visualization as components of value. It is noted that all participants have utilized attacker profiles in their career with positive feedback from clients.

> ”When using attacker profiles I experience that clients begin to visualize threat scenarios and threat agents. This makes an attack yet to occur more real to the client and they understand the importance of a cyberdefence.”

They express that the automated profiling feature in IRCM nulls the time consumption of profiling work, thus encourage the use of this valuable communication channel. They express little concern on the excluding thresholds being calculated guesses at the time as, in their opinion, profiling already is an art of calculated guesses.

Further, in all tests there were expressed appreciation of the structured visualization of an attack that IRCM provides.

> ”The structured, stagewise visualization of the concrete resources and belonging resource alternatives ease the interpretation of an attack. Also, it makes it easy for non cyber-experienced clients to identify resources that can be made

> unavailable. By easing the interpretation of an attack and involving clients in identifying mitigation efforts, IRCM will surely raise the awareness and interest for cyberdefence ”

The claim above was independently supported by three other participants.

The structured coupling of resource trees with kill chain stages is thought to be uniquely useful when building attack scenarios. Answering question 4, five out of eight state that they would use IRCM when building attack scenarios with clients.

> ”Through building attack scenarios together with clients, IRCM facilitates a qualified, customer specific discussion on mitigation efforts ”

The statement above is supported by a consultant with six years of experience in the oil and gas industry. She elaborates that scenario building in IRCM provides a communication platform for industry- and cyber-experts in risk assessment workshops.

> ”Today there is nobody who both fully understand how a SCADA system can be attacked and how all the components that the SCADA system controls work. Through IRCM cybersecurity experts and machine engineers can collaborate on a complementary risk assessment addressing both cybervulnerabilities and the potential physical risk components controlled by a corrupt SCADA system. Such a collaboration will surly reveal unknown risk scenarios”

The remaining three agree that IRCM encourages interdisciplinary collaboration. The stages of the kill chain and the concrete resource alternatives break an attack into smaller parts of tangible components. This is thought to ease the interpretation of an attack as a whole during scenario workshops. With more people gaining knowledge on cyberattacks, the awareness is predicted to increase.

One participant suggests to use the IRCM to build an open source encyclopedia of cyberattacks. Through user contribution a rich catalogue of cyberattacks and their resources and resource alternatives could be built. Such an encyclopedia will allow risk managers to easily compare numerous attacks, their cost, probability and related attacker profiles.

From the future catalogue of attacks a library of common resource alternatives cloud be derived. Using data from previous modelled attacks, IRCM could suggest costs, confidence and identifying attribute values when adding common resource alternatives, e.g. a ransomware-kit or a Raspberry Pi. This is thought to improve the correctness of input values to resource alternatives in addition to speed up the modelling process.

UT2.B

Both experts state that there is a value in modelling a cyberattack from the perspective of what an attacker needs to have, i.e. the required resources, rather than what an attacker need to do. Further, it is pointed out that IRCM is best suited to model a specific attack on a specific target, here a particular vessel.

One of the experts predicts that the derived cost interval can convey a cost-benefit parameter to shipping lines. The surprising low minimal cost of attack is thought to push the cost-benefit measure of cyberdefence in the direction of increased investments.

It is expressed a difficulty on providing an accurate time estimation for the *Motivation* parameter and that the attack as a whole should be associated with a success factor. The success factor should express at what rate an attacker launching the modelled attack will succeed in compromising the target.

Finally, one expert discusses whether the confidence associated with each resource alternative's cost interval is intuitive. It is expressed that it is not intuitive that a large cost interval can be associated with a high confidence, even if the resource alternative is commercially available over the shelf hardware.

7.2.7 UT2.B Validation

Below the resulting Resource Cost Models derived using IRCM during UT2.B are presented.

Back of Bridge Attack

The attack presented by the ECIDS expert from Navtor aims at compromising the navigation system on board a ship in order to harm the vessel. Today's practice involves monthly obliged map updates of the ECDIS. The updates are collected from the map provider, e.g. Navtor, using the online computer at the back bridge. Then the map update is uploaded to an USB-stick that is plugged into the offline, air gaped box at the front bridge running the ECDIS.

If an attacker succeeds in hiding a malware in or corrupting a map update that is uploaded to the air gaped box, the attacker can compromise the ECDIS navigation system. A compromised ECDIS can put its vessel off course and manipulate the perceived speed, resulting in grounding of the vessel or crashes. Consequences of a grounding or a crash are loss of ship and cargo, death of crew or possibly clogging traffic in the case of the target being located in trafficked sea routes, e.g. the Suez or Panama canal.

Figure 7.3 presents the RCM, its derived cost interval and probable attacker profiles of the *Back of Bridge* attack.

Administrative Bridge Network Attack

The *Administrative Bridge Network* attack aims at breaching the bridge network of a vessel before compromising the thruster controller. A successful attack gains the attacker remote, ”hands on the steering wheel” control of the thrusters of the target. This will in practice give the attacker control over the future position of the ship.

The attack requires an extended reconnaissance stage where the adversary builds its own clone of the bridge network on the target vessel in order to scan the network and its components for vulnerabilities. It is required to find vulnerabilities compromising the thruster controller and the network gateway. With a compromised gateway a communication channel to the network can be established in order to control an ongoing attack. To exploit such vulnerabilities the attacker must possess a custom, target specific malware.

As the network is connected online via WiFi inshore, the malware can be delivered using a phishing social engineering scheme. The delivery stage is completed when the phishing email with the malware hidden in an attachment is opened on the administrative bridge network and the attachment is activated. From here the malware should self-install before compromising the thruster controller and the gateway to the administrative bridge network, allowing for a remote satellite communication channel.

With control over the future position of the target vessel an attacker can ground or crash the vessel causing similar consequences as the *Back of Bridge* attack.

Figure 7.4 presents the RCM, its derived cost interval and probable attacker profiles of the *Administrative Bridge Network* attack.

7.2.8 Results

The fact that all 8 test subjects in UT2.A correctly identified and classified resources and belonging resource alternatives, validates the solutions to issue 1, 2 and 3, see section 7.1.7, raised during UT1. Further, it validates that novel users are able to grasp the key concepts of the Resource Cost Model from reading the ”Info”-page.

The only concept being subject for confusion was the meaning of the *Access* attribute. It is unclear whether it describes the required access to the target system, i.e. the intended interpretation, or specific access required to acquire a single resource alternative. Such a specific access could be to be enrolled in a course.

Professional actors in the cybersecurity industry validate the commercial value of IRCM. UT2.A shows that IRCM concretes and structures a cyberattack, making it easier to understand and to identify mitigation efforts. It is recognized that by making all resource alternatives unavailable, the resulting unavailability of the parent resource break the kill chain and thwarts the cyberattack. Should it be impossible to deny all resource alternatives, mitigation efforts must focus on eliminating cheap resource alternatives. By eliminating or increasing the cost of cheap alternatives, the minimum cost of the attack increases, thus lowering the probability of attack.

The monetary cost interval conveys the probability of an attack, and regardless of attacker profile, a cheap attack is considered more probable. Hence, users are able to compare the likelihood of different attacks and focus mitigation efforts thereafter.

The results above yield a mitigation chain identifying how to mitigate the most probable attack. Firstly, by comparing the total cost of attacks, the user identifies the most probable, i.e. cheapest, attack. Secondly, the structured visualization of the kill chain stages and belonging resource trees, proposes resource alternatives to eliminate or to make more expensive to acquire. An unavailable resource mitigates the attack as a whole and an increased cost lowers the attractiveness of the attack.

The automated profiling feature is recognized as useful in personalizing the adversary. By personalizing the adversary, experience show that the defending party start to visualize threat scenarios. This is thought to raise the awareness of the cyberthreat.

The final key value of IRCM confirmed by the participants in UT2.A, is its ability to convey an attack scenario through the resource-based modelling approach. IRCM conveys an attack from A to Z with concrete resources and resource alternatives building a stagewise model showing the workflow of an attack, before deriving the probable attacker profiles. That is, IRCM shows the link from a logic atomic email address and a tangible rangefinder to a Cyber Warrior. This facilitates for discussions between cybersecurity experts understanding the attack as a whole and domain experts on required resources.

Using IRCM, an interdisciplinary team can conclude on a more correct risk assessment and accurate consequence analysis.

The models derived during UT2.B validate that the IRCM installation serves as a vehicle for evaluating a full scale cyberattack in context of the Resource Cost Model.

For the *Back of Bridge* attack IRCM derived the cost interval:

$$[minimum\ cost,\ maximum\ cost] = [\$2300, \$122000] \tag{7.1}$$

with an associated confidence of:

$$confidence = 0.0071604 \tag{7.2}$$

Further, it is estimated to take 75 hours for an attacker to acquire the required resources and realize the attack. The attack requires *Adept* technical skills and some resources must be acquired through *Illegal* actions. Finally, it is only required *External* access to the target system to launch the *Back of Bridge* attack.

For the *Administrative Bridge Network* attack IRCM derived the cost interval:

$$[minimum\ cost,\ maximum\ cost] = [\$7111, \$71520] \tag{7.3}$$

with an associated confidence of:

$$confidence = 0.0000471942 \tag{7.4}$$

Further, the attack is estimated to require a time investment of 55 hours. An attacker must have *Operational* technical skills and be inclined to act *Illegally* to acquire all resources. The *Administrative Bridge Network* attack only requires *External* access to the target system.

The attribute values above derive, for both attacks, the probable attacker profile to be the *Cyber Warrior* profile describing nation state actors with unlimited time, funds and technical skills.

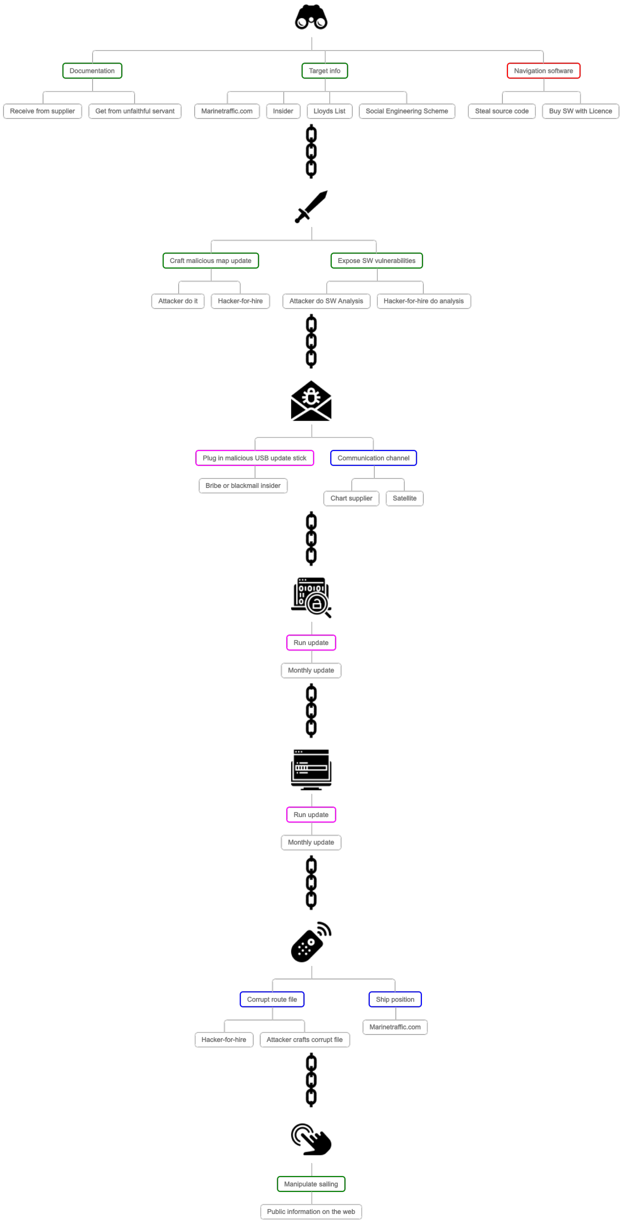


Figure 7.3: Resource Cost Model of the Back of Bridge Attack

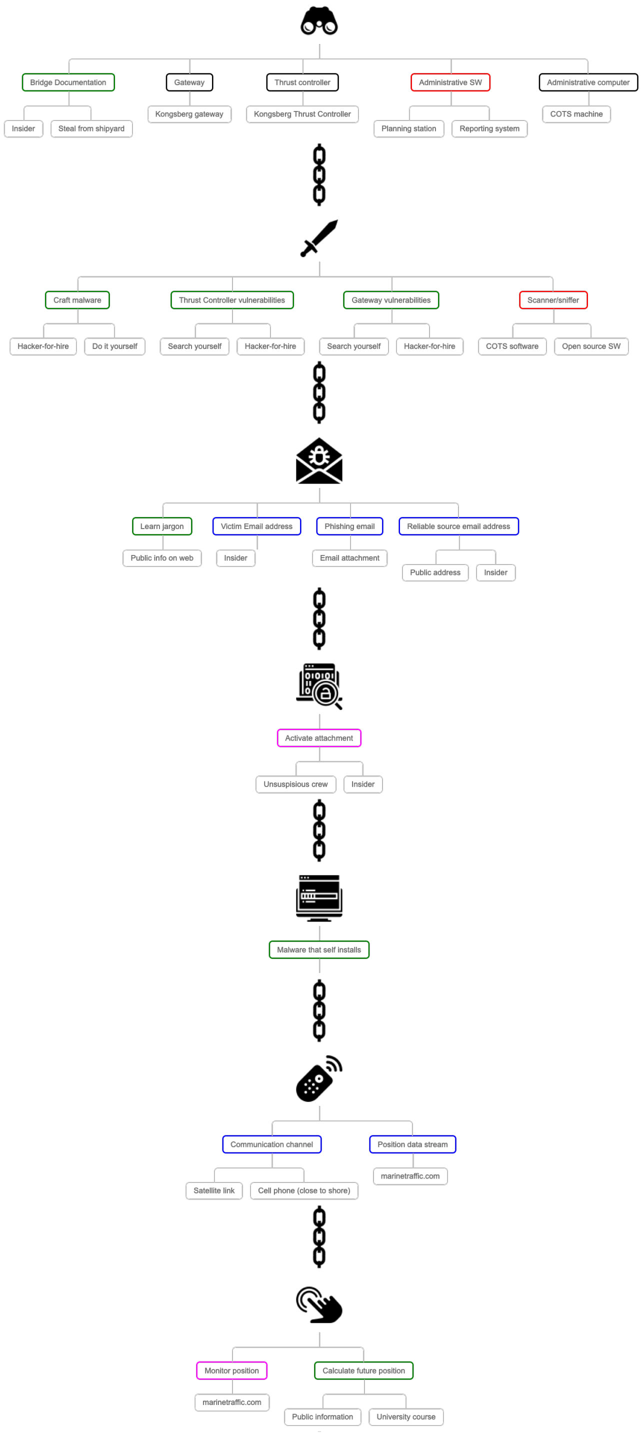


Figure 7.4: Resource Cost Model of the Administrative Bridge Network attack

Chapter 8

Discussion

8.1 Resource Cost Model

8.1.1 Deriving and interpretation of the results

The combination of the Intrusion Kill Chain and attack trees amplifies the advantages and overcomes the shortcomings of the two modelling techniques. The kill chain illustrates the consecutive steps in an attack and communicate how each phase must be completed before the next.

By coupling a resource tree to each stage, RCM generates seven small tree structures. This addresses the issue with graph-based attack models, namely the ability to scale, raised by Hong and Kim (2013). RCM avoids generating large, bewildering attack trees for complex attacks resulting in a conflict between analysis and comprehensibility as discussed by Gadyatskaya and Trujillo-Rasua (2017). A complex tree structure will make the interpretation of the attack as a whole difficult. Also, which resources that are required when and where is unclear in a large, complex attack tree.

In contrast, the seven smaller resource trees explicitly state which resources that are required at the different stages of an attack. This simplifies mitigation efforts, as the user with ease can identify resources that can be made unavailable, thus breaking the kill chain. In addition, the consecutive stages in the kill chain allow the user to identify through which resources an attack can be mitigated at an early stage. Should the defender be unable to break the kill chain, it can pursue to make cheap resource alternatives unavailable. This will increase the minimal cost of the modelled attack, thus make the attack less attractive to the rational adversary.

The *Resource-Resource Alternative* relation illustrate how high-level resources can be realized through different resource alternatives. This relation which communicates how an attack can be made possible through several sets of resource alternatives, is difficult to illustrate in the original kill chain model by Hutchins et al. (2011). It can be argued that a cost estimate derived from the cost of atomic resource alternatives is easier to associate with a high confidence, compared to estimate the cost of high level resources.

Cost Interval

It was observed that the derived cost interval in equation 7.1 is wide with a minimum cost of \$2300 and a maximum estimated cost of \$122 000. It can be argued that a large cost interval does not necessary imply an inaccurate cost estimate, but rather that the evaluated attack can be carried out with a wide span of sophistication and possible impact on the target. Hence, deriving a cost interval, rather than a single cost estimate, provides more confident information on the availability of an attack. An expensive resource set may provide a more stealthy attack with greater impact (a higher success rate) than a cheaper resource set. It will be reasonable to think that the expensive alternative is more attractive to risk averse offenders, not too sensitive to the cost, e.g a national state or a competitor. In this context we relate risk to the probability of being identified as the attacker. On the other hand, the cheap alternative makes the attack available to cost sensitive adversaries willing to take greater risks, e.g. pirates. Should the victim in the aftermath of the attack be able to identify partly or the whole resource set, this would be a significant indicator of who carried out the attack.

Buldas et al. (2006) use attack trees to derive the cost of attacks in order to decide whether the system under protection is a realistic target for gain-oriented attackers. It can be argued that the cost interval serves well as a measurement to compare the likelihood of different cyberattacks. Also, it is reasonable to assume that the cheaper attack is attractive to a larger set of adversaries, and with similar success rate and stealthiness any rational attacker will choose the cheaper alternative. Hence, among a set of similar attacks the cheapest is most likely to incur. This should direct mitigation efforts towards the cheaper, more likely attacks. A subject for further work is to map the minimum cost to a probability scale, e.g. 1-10, in order to ease the process of comparing the likelihood of attacks.

RCM's ability to compare cyberattacks in non-technical, monetary terms through the cost interval, yields a sound argumentation to non-technical actors in order to justify investments in cyberdefence and where to direct mitigation efforts.

The structured visualization and the fact that a single unavailable resource breaks the kill chain, make it possible for non-technical actors to identify mitigation efforts. There is less need for a cyberexpert to identify and make a resource unavailable. An example is to restrict access to the vessel and bridge system documentation required by both the *Back of Bridge* and *Administrative Bridge Network* attack described in 7.2. The non-technical attack representation facilitates for the involvement of actors without cybersecurity background in mitigation efforts. It can be argued that by inviting all fields of knowledge in an organization to participate in a cyberrisk assessment using the IRCM, the resulting risk assessment is more complete and the general cybersecurity awareness in the organization is increased. Further, the structured and easy to comprehend calculation of the cost of launching a cyberattack, communicates the cyberthreat by speaking the lingua franca of money, instead of fear preached in technical terms. With an increased involvement and understanding, follows awareness of the persistent cybersecurity threat yielding increased mitigation efforts.

Confidence of cost interval

The derived confidences associated with the *Back of Bridge* and *Administrative Bridge Network* attack, shown in equation 7.2 and 7.4 are low. From equation 5.5 it follows that the total confidence associated with a cost interval becomes low with a significant number of resources. The low confidence value follows regardless of all resource alternatives being associated with a high confidence. It can be argued that it is natural that a greater number of resources yields greater uncertainty associated with the total cost, although the confidence associated with each resource is high.

Further, it can be nonintuitive that a wide cost interval can be assessed a high confidence. An example being the cost interval of a commercially over the shelf (COTS) available laptop, which has a price range, i.e. cost interval, from USD \$100 to \$10,000. Here the cost interval is wide, but one can be certain that a COTS laptop cost no less than \$100 and that the price rarely exceeds \$10,000, thus the cost interval is given a high confidence.

The main benefit of the confidence is for attack comparison and the derived number does not yield any sort of probability on an actual attack cost being within the derived cost interval.

The derived confidence associated with the *Administrative Bridge Network* attack is negative two orders of magnitude of ten lower than the confidence associated with the *Back of Bridge* attack. This is partly due to the fact that the former attack requires 19 resources contributing to uncertainty compared to the latter attack which only requires 12. Both attacks make use of bribing or black mailing insiders and hacking-for-hire which yield a low confidence. Further, none of the attacks only require common commercially available hardware which would have yielded a higher confidence. With custom software, special hardware and use of insiders, it is reasonable to accept a low confidence in the derived cost intervals.

To revise the calculation of the confidence such that it yields more insight than just a number for comparison, is subject for further work. An improved formulation could associate a probability distribution to the cost interval of each resource alternative. E.g. a resource alternative with a minimum cost of \$10 and a maximum cost of \$100 could be associated with a distribution showing a 90% probability that the cost is between \$10-50 and 10% probability that the cost is between \$50-100. Further, the distributions of all resource alternatives can be aggregated to yield a distribution for the total cost interval of the whole attack.

Assigning cost and confidence

The task of assigning a minimal and maximal cost to a resource alternative before associating a confidence with the costs, can be a difficult task. In the case of common COTS available products it is easy to find a cost interval associated with a high confidence. In contrast, bribing or blackmailing insiders and custom malware only available through the Darkweb are resource alternatives which it is difficult to assign a cost. Although there is a wide span of resources in all five resource classes - *Skill*, *Tangible*, *Logic*, *Logic-atomic* and *Behavioral* - it can be argued that there are some commonalities within resource classes.

Skill - The cost of gaining domain knowledge can often have a wide cost interval associated with a high confidence. This follows from the fact that there is free information on the web on almost any subject, while the cost of taking a university course can be high. Despite the difference in cost of the two generic alternatives, both costs are associated with a high confidence as the prices are zero or publicly available.

In the matter of an attack requiring custom software, Boehm et al. (2000) survey paper on software development cost estimation techniques concludes that there is no single technique best suited for all development situations. This implies that a cost estimate on a custom software should be associated with a low confidence, and can be observed in both attacks from UT2.B described in section 7.2.7. Here, the costs of hiring hackers to build a custom, malicious map update and a malware to breach the bridge network are associated with a confidence of 0.2 and 0.3. Note that the cost of the resource alternative of the attacker crafting the malware itself is associated with a cost of 0 and a confidence of 1, as this only requires an investment in time by the attacker which is measured by the *Motivation* attribute.

A subject for further work is to implement a software development cost estimation feature into the IRCM. Such a feature could be based on the required skill level and an estimate of the number of code lines required. This would be inaccurate and should be associated with a low confidence, but it can be argued that a goal of RCM is to identify the significant costs of an attack. Should an attack require a large custom software this will contribute greatly to the total cost, while the accuracy of the estimate is of less importance.

Tangible - Being mostly COTS available hardware, e.g. micro controllers, laptops or antennas, the cost interval of tangible resources can be associated with a high confidence. However, there exist attacks requiring illegal hardware such as a military grade GPS Jammer for the GNSS spoofing attack presented in UT1 and UT2.A, see Appendix C.1 and C.2. Hardware and other tangible resources only available on the illegal marked will be more difficult to assign a cost interval, thus yielding a lower confidence.

Logic - Freely and COTS available software can easily be assigned a cost interval with a high confidence. In contrast, Meland and Sindre (2019) show that the price of hackertools and malware-kits on the Darkweb varies and Meland et al. (2020) state that ransomware-kits often are fakes. From this it follows a low confidence associated with logic resources acquired on the illegal market.

Logic-atomic - Email addresses, passwords and other logic-atomic resources must in many cases be acquired through auxiliary cyberattacks such as phishing or other social engineering schemes. Hence, the same difficulty of assigning a cost and confidence to illegal logic resource alternatives yields for many logic atomic resource alternatives. As an example, the *Administrative Bridge Network* attack requires a malicious email with the malware attached to be delivered to the target bridge network. This requires an address to a mailbox which is opened on the bridge network and a phishing scheme with an unsuspicious email from a reliable source. The three resources are associated with a confidence of 0.3, 0.1 and 0.3.

Behavioral - The cost interval of resource alternatives in this category may be difficult to estimate with any confidence. As in the *Back of Bridge* attack in UT2.B, the resource alternative to bribe or blackmail a crew member to insert a USB-drive on the bridge is hard to assign a cost interval. An approach to estimate the cost of a bribe is to use the salary of the crew member, but the integrity of the crew member and its loyalty to its employer are

”prize-less” factors that will affect the cost.

Resource alternatives relying on human errors, e.g. opening a malicious email from a phishing attack, can be assigned a cost of 0 with a confidence of 1 as RCM models the attacker’s cost of launching a cyberattack. A sophisticated, more expensive phishing attack will have a higher success rate, but the final action, the behavioral resource, of the victim completing the attack has no cost to the attacker.

8.1.2 Validity and limitations

The results from UT2.B in section 7.2.7 presenting the two full scale attack scenarios *Back of Bridge* and *Administrative Bridge Network* attack validate that RCM is able to model a cyberattack based on the required resources. This answers research question 1: How can the cost of launching a cyberattack be estimated? The two modelled attack scenarios from the maritime domain serves only as a proof of concept and further work should include to validate RCM’s ability to model attack scenarios in other domains.

Broadhurst et al. (2018) observed an increase in interaction between cybercriminals and state or quasi-state cybersecurity actors on Darkweb hacker markets. These interactions often involve offensive cyberoperations. As a result, the sophistication and value of malware and hackertools available to cybercriminals have increased. This include zero-day exploits. As RCM is a tool to model known cyberattacks, it falls short of modelling zero-day exploits. Hence, there is a persistent cyberthreat even to organisations being able to make at least one resource unavailable in each known, RCM modelled attack. Zero-day exploits may also repair a kill chain thought to be broken, thus any responsible organisation should strive to make as many resources as possible unavailable in addition to constantly monitoring its network traffic to identify erupting attacks.

8.2 Attacker Profiling

The purely inductive criminal profiling methodology presented in chapter 6 is not extensive nor satisfactory. To argue that a profiling methodology based on five parameters that do not fully explain attacker motivation and do not make use of forensic data would be highly speculative. However, it can be argued that the chosen identifying attributes - *Motivation*, *Technical Skills*, *Cost*, *Limit* and *Access* - are closely linked to a criminal’s *Motive*, *Means* and *Opportunity*. This trinity makes up the MMO-profile of a crime, see section 3.3.1, which is widely used in criminal investigations and law enforcement.

The investment in time described by the *Motivation* attribute, creates or limits the *Opportunity* of attack, i.e. circumstances that make a crime possible. The more time an attacker is motivated to invest, the more methods of attack and resources are available. Hence, a larger investment in time results in greater opportunities.

Both *Technical Skills* and *Cost* are attributes related to criminal’s *Means*, i.e. the instruments available to the attacker. In cybercrime this includes hacking tools, attack methods, technology used etc. All of these are limited by an attacker’s technical skills or financial capacities, thus may be used to exclude cybercriminal profiles from the pool of probable attackers. In addition, the *Cost* attribute may constrain the *Opportunity* of a criminal profile to launch a certain cyberattack.

The *Limits* of an attacker describes its moral *Means*. That is, does the moral limits of the attacker align with the required ethics to carry out the crime. With the moral ability to break the law, it follows new opportunities as new attacks are morally available.

Finally, the *Access* level of a hostile actor directly sets its *Opportunities*. With an internal access level, the possible targets within the system and attack methods available to the attacker increase.

It is identified that none of the five identifying attributes have the ability to touch the *Motive* of an MMO-profile. *Motive* describes why a rational agent commit a crime, e.g. financial gain or revenge, and is empathized by the literature reviewed in chapter 4 as a key identifying attribute of an attacker profile. This shows that the methodology is not accurate enough to conclude on a single probable cybercriminal profile. However, it can be argued that by covering two out of three elements of an MMO-profile the profiling methodology can confidently exclude profiles from the set of probable attackers.

Routine Activities Theory, see section 3.3.2, states the required conditions for a rational cyberattack to occur. The conditions are: There exist 1) an accessible and attractive target, 2) the absence of a capable guardian and the presence of 3) a motivated offender with 4) the resources required to commit the crime. It can be argued that only the two latter conditions are controlled by the attacker, thus should be accounted for in a profiling methodology. The required time investment described with the *Motivation* attribute shows a motivated offender and the remaining attributes identify if a profile possesses the required resources.

With the identifying attributes relating to aspects in both MMO-profiling and Routine Activities Theory, it can be argued that the profiling methodology is well anchored in the study of criminal behavioral. Still, it is recognized that the attributes are not satisfactory for a sound profiling methodology, but can confidently identify improbable attacker profiles.

The study finds it reasonable that both attacks from UT2.B suggest *Cyber Warrior* as the only probable attacker profile. The minimal costs of the attacks are USD $2300 and $7111, which exclude cost sensitive profiles such as Script Kiddies, Hacktivists, Cyber Vandals and Internals. Further, the large time investment of 75 and 55 hours exclude Petty Criminals looking for an easy attack with a bang for the buck and terrorists are excluded for their lack in technical skills. The *Back of Bridge* attacks requires adept technical skills, thus excluding Mobsters from the set of probable attacker profiles. For the *Administrative Bridge Network* attack, mobsters are excluded by the high minimal cost and the large time investment. It is empathized that further work will be to validate the exclusion rules through empirical data, as the numerical thresholds for motivation and cost is only a guess at the time. One approach is to do a case study on known cyberattacks where the attacker was apprehended.

8.3 IRCM

The derived models presented in section 7.2.7 validate that the IRCM installation serves as a vehicle for evaluating a full scale cyberattack in the Resource Cost Model, hence answer research question 1: How can the cost of launching a cyberattack be estimated? Further, UT1, UT2.A and UT2.B validate that IRCM conveys the key concepts of RCM to novel users such that they are able to utilize IRCM to build models of their own. The exception being the *Access* attribute which was subject to some confusion. To review the presentation of this attribute is highly prioritized work in order to secure that IRCM is correctly used.

The IRCM artifacts show that an implementation of the cybercriminal profile exclusion rules derived in chapter 6 are able to derive a set of probable attacker profiles. This answers research question 2: Which rational agents of attack do the required resources and the cost of a cyber-attack imply? Further, actors in the cybersecurity community find the profile feature to be of value. It is recognized as useful in personalizing the adversary and initiates a visualization of the adversary and attack scenarios. It can be argued that this visualization process induces a raised awareness of the persistent cyberthreat.

The study finds IRCM to provide an interactive framework to construct threat scenarios. This makes IRCM an excellent tool to drive interdisciplinary risk assessment workshops. Interviews with cybersecurity consultants from Mnemonic AS with experience from the petroleum industry suggest that the interdisciplinary collaboration will yield more accurate risk assessments and attack scenarios.

> ”Today there is nobody who both fully understand how a SCADA system can be attacked and how all the components that the SCADA system controls work. Through IRCM cybersecurity experts and machine engineers can collaborate on a complementary risk assessment, addressing both cybervulnerabilities and the potential physical risk components controlled by a corrupt SCADA system. Such a collaboration will surly reveal unknown risk scenarios”

Further, IRCM encourages engagement of non-cybersecurity clients in threat scenario building with cybersecurity consultants. Increased client engagement will induce more client specific, accurate threat scenarios. The fact that IRCM's design allowed all test subjects to interpret the main concepts of RCM and construct models of their own within 15 minutes, gives confidence to state that clients with various background can easily be engaged in scenario construction. Finally, the study argues that engagement in scenario construction will drive an increased cyberthreat awareness.

The low minimum costs derived when validating IRCM during UT2.B with experts in maritime, illustrate the significant bias in cost between carrying out a cyberattack and the resulting consequences. As an example, AIBN and DAIBN (2019) description of the events leading up to the KNM Helge Ingstad incident suggests that a *Back of Bridge* attack could have similar consequences. The bias is obvious from the estimated minimum attack cost of USD $1400 compared to the USD $475 million expense of the KNM Helge Ingstad incident. It can be argued that this bias provides a rational incentive of attack. With ransomware-kits available for USD $44 (Broadhurst et al., 2018) and the fact that the WannaCry attack on Møller Mærsk caused financial losses in excess of $400 million (Moller-Maersk, 2017), the argumentation holds for attacks outside the maritime domain as well.

With the persistent cyberthreat from zero-day exploits and the rational incentive of attack elaborated above, any user of IRCM should strive to make as many resources as possible unavailable. The RCMs derived in UT2.B revile that many of the required resources are not controlled by the target. Hence, making it unrealistic to make such resources unavailable. Such resources include university courses teaching required skills, thrust controllers and other third party hardware. Nevertheless, the study speculates that there will always be at least one resource controlled by the target that can be made unavailable, thus mitigating the entire attack. In both attacks in UT2.B such a target-resource should be vessel and bridge network documentation. Should no resource be controlled by the target, a mitigation strategy is to remove cheap resource alternatives in order to drive the minimum cost of an attack up. The profiling methodology from chapter 6 states that a higher cost will eliminate cybercriminal profiles from the set of probable attackers.

From the RCMs modelling the *Back of Bridge* and *Administrative Bridge Network* attack, it is observed that most of the resources are required in the early stages of the kill chain. In contrast the later stages - *Exploitation*, *Installation*, *Command and Control* and *Actions on Objective* - require few resources. This supports the general criticism towards the cyber kill chain that it focuses too much on the perimeter and malware attack vector (Pols, 2017). Therefore, future improvements could be to include other sets of stages more suitable to describe attacks for instance related to social engineering, denial-of-service or code injection as suggested by Hospelhorn (2020). Pols (2017) suggests a 18 stage kill chain by adding stages from the MITRE ATT&CK framework to the original cyber kill chain. 18 stages are more stages than there are resources in the *Back of Bridge* attack. Hence, a middle ground could be to start with a longer, extensive scaffold kill chain and let users delete unnecessary stages.

The observed crowded resource trees in the early stages of the two specific attacks in UT2.B, suggest that a model of a more generic attack will result in large, over-crowded resource trees. When IRCM is used on a small, e.g. laptop, screen the over-crowded resource trees might ”break” the CSS dictating the structured visualization. Any broken CSS will not cause the application to crash or loose any functionality, but one of its strong suites, the structured visualization, suffers. Based on the fact that the target users of IRCM are professional actors usually working from a working station with a larger monitor, it can be argued that this issue will mostly go unrecognized. However, best practice for any web application is for it to function on all screen sizes.

The above suggests to encourage users to model more specific attacks, rather than generic attacks. This is supported by the fact that a more target specific attack allows the user to recognize more target specific resources, thus deriving a more accurate model. Using the *Administrative Bridge Network* attack which requires vessel and bridge system documentation, a target specific model, i.e. a model for a specific vessel, will have more accurate information on the availability of the documentation. In some cases the vessel documentation might be publicly available, yielding an obvious mitigation effort. For another target, the documentation might be available at the origin shipyard and the current owner of the ship must contact the shipyard in order to make the documentation unavailable. With more target specific models, follows target specific mitigation efforts.

Target specific models may induce tedious, repetitive work for an organization with many potential targets. In order ease the modelling work, future IRCM versions should save and aggregate similar, commonly used resources and resource alternatives in a library. For example, for any attack requiring any sort of documentation, the ”Bribe insider” alternative is an option. With a working library feature, IRCM could automatically suggest attribute values - minimum and maximum cost, confidence, motivation, skill, legal limit and access - based on previous declarations of the ”Bribe insider” alternative. Such peer reviewed alternative attribute values can be argued to be more accurate than the calculated guesses of a single consultant.

With a wide use of IRCM with a library feature, the potential is an encyclopedia of peer reviewed resources and resource alternatives building a catalog of target specific attacks. A future IRCM encyclopedia can be a compliment to MITRE CVE (MITRE, 2020). Predetermined resources and resource alternatives will induce a rapid RCM construction and provide peer reviewed attribute values to common resource alternatives. To suggest peer reviewed attribute values may overcome the difficulty and uncertainty users express when estimating cost and required time to realize a resource alternative. Further, a catalog of target specific attacks lets users compare their specific model to similar targets. Such a comparison might revile left out resource alternatives, potentially making thought of unavailable resources available. From this it can be argued that a library feature will increase the accuracy and extensiveness of models.

8.3.1 Method error

The study incorporated user testing with a single participant in each test as the only method for usability inspection. Grondin et al. (2002) suggest that paired-user testing allows a better understanding of conceptual misunderstandings and difficulties, while individual tests better identify usability problems. This suggests that to evaluate the ”info”-page which conveys the concepts of RCM, would yield a more sound evaluation with paired testing. Tan et al. (2009) state that user testing is better suited to test the whole user scenario compared to heuristic analysis. Heuristic analysis relies on experts evaluating a product’s design in the context of predefined design principles and tends to cover more high-level structural problems. In a heuristic analysis the evaluators assess every aspect of the product, while user tests will only assess the features addressed in the predefined scenario. If the scenario does not include a possibility of failure or misuse of a feature, then problems remain salient from user testing. Tan et al. (2009) conclude that user testing and heuristic analysis are complimentary and both techniques of usability evaluation are needed to achieve an optimal design and user experience. It is recognized that paired-user testing and heuristic analysis would have complimented our usability inspection of IRCM.

The study used different user groups for all tests. This induced a higher total number of test subjects, but denied test subjects to observe the development of the artifacts. A higher number of participants validating the usability implies a higher confidence of the usability of the artifacts in general populations. More diverse test groups would also boost confidence in usability. With only test subjects with a general cybersecurity understanding and an above average technical skill set, the test group’s diversity is subject to improvement. In a survey paper Bastien (2010) concludes that the question of the number of users to test is far from being solved and requires further research.

The study recognizes the issue of convenience sampling when finding participants for user testing. Firstly the participants were picked due to their knowledge of the cyberdomain, but their accessibility as colleagues contributed to their participation. Etikan et al. (2016) state that an assumption for justifying convenience sampling is that there would be no difference in the research results obtained from a random sample, a nearby sample, a co-operative sample, or a sample gathered in some inaccessible part of the population. Further, a weakness of the study rooted in the selection of user test subjects, is that no test subjects participated in all rounds of testing. This would allow for the subject to observe the development of the product and provide feedback on whether new features and design choices yielded an improvement from previous versions.

When gathering qualitative data from user test observations and debrief interviews, there is a considerable observer bias. The research recognizes that observer bias will have affected which observations that are empathized and researchers might unconsciously have cherry picked observations that fit their suspicions and hypothesis. Hence, user tests conducted independently of the research team developing RCM and IRCM provide stronger validation of the model and its installation. Further, it is recognized that the fact that product developers also conducted testing and that the subjects were aware of this, may have induced a higher threshold for product criticisms.

To further substantiate the modelling capabilities and the value of RCM, the study could have compared the results from UT2.B with results derived from modelling the same attack scenarios in other cyberattack modelling frameworks such as MITRE ATT&CK. Strom et al. (2018) and Strom et al. (2017) present the MITRE ATT&CK framework, a behavioral-based threat model, to increase visibility and improve the communication of how effective security efforts across an organization are. The study argues that the ATT&CK behavioral-based threat model and other action based models are to different from the RCM resource based threat model for any comparison to be of value.

8.4 Further work

From the discussion above the following further work is identified and presented in prioritized order:

1. Clarify that the *Access* profile identifying attribute describes the required access level to the target system.

2. Case study on known cyberattacks to set more accurate thresholds for the exclusion rules used for cybercriminal profiling.

3. Map minimum cost of an attack to a probability scale in order to ease comparison of the probability of different attacks.

4. Let users chose different kill chain models and allow users to delete stages without relevance to the current modelled attack.

5. Provide a library of common resource alternatives. The alternatives in the library should have predetermined costs, confidence and profile identifying attribute values.

6. Make an application structure allowing for users to share, comment and revise each other's models. This should facilitate for users to contribute to an open source encyclopedia of RCM modelled cyberattacks and their required resources and resource alternatives.

7. As the cheapest set of resource alternatives realizing an attack is deemed the most attractive to any attacker, IRCM could highlight these resource alternatives. This would ease the identification of which resource alternatives the defending party

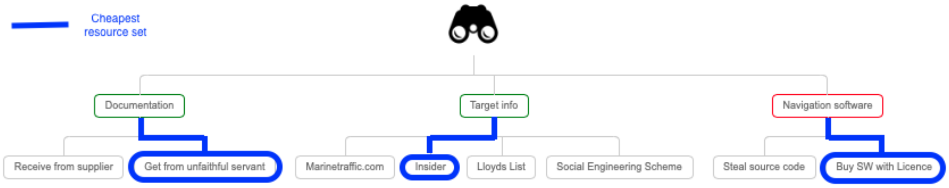


Figure 8.1: Sketched solution to highlight cheapest resource alternatives

should prioritize to make unavailable. A sketched solution for this feature is shown in Figure 8.1.

8. Implement a gamification of the IRCM as an addition to the current modelling tool. A gamification could be a story-game where the users ”investigate” a cyberasset. Through the investigation of the assets users should recognize required resources and resource alternatives. A user will win the game if all resources are identified and the model is correctly constructed.

A short paper based on work from this thesis was peer-reviewed and accepted for ”The Seventh International Workshop on Graphical Models for Security (June 22, 2020)” (GraMSec, 2020) and it will be published in Lecture Notes in Computer Science series by Springer. A pre-print of the short paper can be viewed in Appendix A.

Chapter 9

Conclusion

Through the iterative nature of design science the study has continuously improved the RCM and its installation IRCM. The resource cost modelling approach conveys an attack from A to Z with concrete resources and resource alternatives. IRCM builds a stagewise model showing the workflow of an attack, before it derives the probable attacker profiles based on the required resources. That is, the IRCM tool shows the link from an email address, a rangefinder and a ransomware-kit to a Cyber Warrior. Further, the structured visualization of this link serves as a platform for interdisciplinary cyberthreat scenario assessments.

The RCM and IRCM have been validated in a maritime context. However, both are considered to be work in progress with many potential improvements related to usefulness and usability. To ensure that the artefacts could have a wider usage than just the maritime context, further user testing and evaluation is needed.

Nevertheless, there is no silver bullet to threat modelling. The study is trying to address a real-world problem of missing historical incident data, which is a particular concern for new technology. The RCM provides best accuracy with specific attacks on a certain target; when there are few resources and resource alternatives. Hence, the RCM modelling approach is not recommended to represent attacks with several possible attack vectors of different types. In such cases, multiple RCMs could be created in parallel and be compared. To generate such a set of RCMs to represent a generic, target unspecific attack, quickly becomes a tedious task. As always, the analyst should choose the right tool for the job at hand. RCM is a hammer, but not all cyberattacks are nails.

Bibliography

Ablon, L., Libicki, M.C., Golay, A.A., 2014. Markets for cybercrime tools and stolen data: Hackers' bazaar. Rand Corporation.

AGCS, 2018. Allianz Global Corporate and Speciality SE, Safety and shipping review 2018. Technical Report 2018. Allianz Global Corporate & Specialty SE. Fritz-Schaeffer-Strasse 9, 81737 Munich, Germany.

AGCS, 2019. Allianz Global Corporate and Speciality SE, Safety and shipping review 2019. Technical Report 2019. Allianz Global Corporate & Specialty SE. Fritz-Schaeffer-Strasse 9, 81737 Munich, Germany.

AGCS, 2020. Allianz Risk Barometer 2020. Technical Report. Fritz-Schaeffer-Strasse 9, 81737 Munich, Germany. `https://www.agcs.allianz.com/content/dam/onemarketing/agcs/agcs/reports/Allianz-Risk-Barometer-2020.pdf`.

AIBN, DAIBN, 2019. Part One Report on the Collision on 8 November 2018 Between the Frigate HNoMS HELGE INGSTAD and the Oil Tanker SOLA TS Outside the STURE Terminal in the Hjeltefjord in Hordaland County. Technical Report. Accident Investigation Board Norway, Defence Accident Investigation Board Norway. P.O. Box 213, N-2001 Lillestrøm, Norway. `https://www.aibn.no/Sjofart/Avgitte-rapporter/2019-08-eng`.

Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., Durumeric, Z., Halderman, J.A., Invernizzi, L., Kallitsis, M., et al., 2017. Understanding the mirai botnet, in: 26th {USENIX} Security Symposium ({USENIX} Security 17), pp. 1093–1110.

Arntz, P., 2018. How threat actors are using SMB vulnerabilities. Malwarebytes Lab. Malwarebytes, 3979 Freedom Circle, 12th Floor Santa Clara, California 95054 U.S. `https://blog.malwarebytes.com/101/2018/12/how-threat-actors-are-using-smb-vulnerabilities/`.

Assante, M.J., Lee, R.M., 2015. The industrial control system cyber kill chain. SANS Institute InfoSec Reading Room 1.

Bagnato, A., Kordy, B., Meland, P.H., Schweitzer, P., 2012. Attribute decoration of attack–defense trees. International Journal of Secure Software Engineering (IJSSE) 3, 1–35.

Bastien, J.C., 2010. Usability testing: a review of some methodological and technical aspects of the method. International journal of medical informatics 79, e18–e23.

Boehm, B., Abts, C., Chulani, S., 2000. Software development cost estimation approaches—a survey. Annals of software engineering 10, 177–205.

Borges, M.A., Stepnowsky, M.A., Holt, L.H., 1977. Recall and recognition of words and pictures by adults and children. Bulletin of the Psychonomic Society 9, 113–114.

Boyes, H., Isbell, R., 2017. Code of Practice - Cyber Security for Ships. Technical Report. Institution of Engineering and Technology, London, United Kingdom. URL: `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/642598/cyber-security-code-of-practice-for-ships.pdf`. [Online; accessed 19-May-2020].

Broadhurst, R., Lord, D., Maxim, D., Woodford-Smith, H., Johnston, C., Chung, H.W., Carroll, S., Trivedi, H., Sabol, B., 2018. Malware trends on 'darknet'crypto-markets: Research review. Available at SSRN 3226758 .

Brown, T., Katz, B., 2011. Change by design. Journal of product innovation management 28, 381–383.

Buldas, A., Laud, P., Priisalu, J., Saarepera, M., Willemson, J., 2006. Rational choice of security measures via multi-parameter attack trees, in: International Workshop on Critical Information Infrastructures Security, Springer. pp. 235–248.

Casey, T., 2007. Threat agent library helps identify information security risks. Intel White Paper 2.

Casey, T., 2015. Understanding cyber threat motivations to improve defense. Intel White Paper .

Chen, Q., Bridges, R.A., 2017. Automated behavioral analysis of malware: A case study of wannacry ransomware, in: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE. pp. 454–460.

CNNSI, 2015. Committee on National Security Systems Glossary, CNSSI No. 4009. Committee on National Security Systems. CNSS Secretariat National Security Agency, 9800 Savage Road, Ste 6165 Fort George G. Meade, Maryland 20755-6716, U.S. `http://www.cnss.gov/CNSS/openDoc.cfm?EfAwzHrSbR3CL0Lmg/4FKA==`.

Cohoen, L.E., Felson, M., 1979. Social change and crime rate trends: A routine activity approach. American Sociological Review 44, 588–608.

Hasso Plattner Institute of Design, a.S., 2020a. https://static1.squarespace.com/static/57c6b79629687fde090a0fdd/t/58ac891ae4fcb50f1fb2f1ab/1487702304601/Facilitator%27s+Guide_Design+Thinking.pdf.

Hasso Plattner Institute of Design, a.S., 2020b. https://dschool.stanford.edu/resources/a-virtual-crash-course-in-design-thinking.

Diogenes, Y., Ozkaya, E., 2018. Cybersecurity??? Attack and Defense Strategies: Infrastructure security with Red Team and Blue Team tactics. Packt Publishing Ltd.

Ekblom, P., Tiley, N., 2000. Going equipped. The British Journal of Criminology 40, 376–398.

Etikan, I., Musa, S.A., Alkassim, R.S., 2016. Comparison of convenience sampling and purposive sampling. American journal of theoretical and applied statistics 5, 1–4.

Europol, 2017. Serious and Organized Crime Threat Assessment 2017. Technical Report. Europol. Europol, P.O. Box 908 50, 2509 LW The Hague, The Netherlands. https://www.europol.europa.eu/activities-services/main-reports/european-union-serious-and-organised-crime-threat-assessment-2017.

Europol, 2018. Internet Organized Crime Threat Assessment 2018. Technical Report. Europol. Europol, P.O. Box 908 50, 2509 LW The Hague, The Netherlands. URL: https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment-iocta-2018.

Europol, 2019. Internet Organized Crime Threat Assessment 2019. Technical Report. Europol. Europol, P.O. Box 908 50, 2509 LW The Hague, The Netherlands. URL: https://www.europol.europa.eu/iocta-report. [Online; accessed 19-May-2020].

Gadyatskaya, O., Trujillo-Rasua, R., 2017. New directions in attack tree research: Catching up with industrial needs, in: International Workshop on Graphical Models for Security, Springer. pp. 115–126.

Goldkuhl, G., 2012. Pragmatism vs interpretivism in qualitative information systems research. European journal of information systems 21, 135–146.

Grabosky, P.N., 2001. Virtual criminality: Old wine in new bottles? Social and Legal Studies 10, 243–249.

GraMSec, 2020. The Seventh International Workshop on Graphical Models for Security. Augusta University and Leiden University and University of Luxembourg. URL: https://gramsec.uni.lu.

Greenberg, A., 2018. The Untold Story of NotPetya, the Most Devastating Cyberattack in History. Wired. URL: https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/.

Grondin, N., Bastien, J.C., Agopian, B., 2002. Les tests utilisateurs: avantages et inconvénients des passations individuelles et par paires, in: Proceedings of the 14th Conference on l'Interaction Homme-Machine, pp. 121–128.

Hadnagy, C., 2010. Social engineering: The art of human hacking. John Wiley & Sons.

Haga, K., 2019. Cybercrime Economy. URL: http://folk.ntnu.no/kristaha/Prosjektoppgave.pdf.

Hemenway, K., 1982. Psychological issues in the use of icons in command menus, in: Proceedings of the 1982 conference on Human factors in computing systems, pp. 20–23.

Hevner, A., Chatterjee, S., 2010. Design science research in information systems, in: Design research in information systems. Springer, pp. 9–22.

Hong, J.B., Kim, D.S., 2013. Performance analysis of scalable attack representation models, in: IFIP International Information Security Conference, Springer. pp. 330–343.

Hospelhorn, S., 2020. What is The Cyber Kill Chain and How to Use it Effectively. Technical Report. Varnois. URL: https://www.varonis.com/blog/cyber-kill-chain/. [Online; accessed 12-May-2020].

Hutchins, E.M., Cloppert, M.J., Amin, R.M., 2011. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. Leading Issues in Information Warfare & Security Research 1, 80.

Hydro, 2019. Cyber-attack on Hydro. URL: https://www.hydro.com/en/media/on-the-agenda/cyber-attack/. [Online; accessed 18-May-2020].

Jahankhani, H., Al-Nemrat, A., 2012. Examination of cyber-criminal behaviour. International Journal of Information Science and Management (IJISM) , 41–48.

Jensen, P.G., Larsen, K., Legay, A., Poulsen, D., 2017. Quantitative evaluation of attack defense trees using stochastic timed automata, in: International Workshop on Graphical Models for Security, HAL Id: hal-01640091. pp. 75–90.

Jordan, B., Piazza, R., Wounder, J., 2017. STIX Version 2.0. Part 1: STIX Core Concepts. Technical Report. OASIS Commettee Specifications 01. http://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part1-stix-core.html.

Kirwan, G., Power, A., 2012. The psychology of cyber crime: Concepts and principles. Evolution 2021, 199.

Kjaerland, M., 2006. A taxonomy and comparison of computer security incidents from the commercial and government sectors. Computers & Security 25, 522–538.

Knight, E., Gunawardena, C.N., Aydin, C.H., 2009. Cultural interpretations of the visual meaning of icons and images used in north american web design. Educational Media International 46, 17–35.

Kohlberg, L., 1974. The claim to moral adequacy of a highest stage of moral judgment. The journal of philosophy 70, 630–646.

Kordy, B., Piètre-Cambacédès, L., Schweitzer, P., 2014. Dag-based attack and defense modeling: Don't miss the forest for the attack trees. Computer science review 13, 1–38.

Kshetri, N., 2006. The simple economics of cybercrimes. IEEE Security & Privacy 4, 33–39.

Kumar, R., Ruijters, E., Stoelinga, M., 2015. Quantitative attack tree analysis via priced timed automata, in: International Conference on Formal Modeling and Analysis of Timed Systems, Springer. pp. 156–171.

Kwan, L., Ray, P., Stephens, G., 2008. Towards a methodology for profiling cyber criminals, in: Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008), IEEE. pp. 264–264.

Lee Badger, Murugiah Souppaya, M.T.E.T.D.Y.K.S., 2016. NIST Special Publication 800-179: Guide to Securing Apple OS X 10.10 Systems for IT Professionals: A NIST Security Configuration Checklist. The National Institute of Standards and Technology. 100 Bureau Drive, Gaithersburg, Maryland 20899, U.S. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-179.pdf.

Liska, A., Gallo, T., 2016. Ransomware: Defending against digital extortion. " O'Reilly Media, Inc.".

Manky, D., 2013. Cybercrime as a service: a very modern business. Computer Fraud & Security 2013, 9–13.

March, S.T., Smith, G.F., 1995. Design and natural science research on information technology. Decision support systems 15, 251–266.

McGuire, M., 2012. Organised crime in the digital age. London: John Grieve Centre for Policing and Security .

McKendall, M.A., Wagner III, J.A., 1997. Motive, opportunity, choice, and corporate illegality. Organization Science 8, 624–647.

McQueen, M.A., Boyer, W.F., Flynn, M.A., Beitel, G.A., 2006. Quantitative cyber risk reduction estimation methodology for a small scada control system, in: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), IEEE. pp. 226–226.

Meadows, C., 1998. A representation of protocol attacks for risk assessment, in: Proceedings of the DIMACS Workshop on Network Threats, pp. 1–10.

Meland, P.H., Bayoumy, Y.F.F., Sindre, G., 2020. The ransomware-as-a-service economy within the darknet. Computers & Security , 101762.

Meland, P.H., Sindre, G., 2019. Cyber attacks for sale .

Microsoft, 2020. https://www.microsoft.com/en-us/msrc/bounty.

MITRE, 2019a. CAPEC-89: Pharming. The MITRE Corporation. 202 Burlington Road Bedford, MA 01730-1420, Massachusetts , U.S. https://capec.mitre.org/data/definitions/89.html.

MITRE, 2019b. CAPEC-94: Man in the Middle Attack. The MITRE Corporation. 202 Burlington Road Bedford, MA 01730-1420, Massachusetts , U.S. https://capec.mitre.org/data/definitions/89.html.

MITRE, 2019c. CAPEC-98: Phishing. The MITRE Corporation. 202 Burlington Road Bedford, MA 01730-1420, Massachusetts , U.S. https://capec.mitre.org/data/definitions/98.html.

MITRE, 2020. Common Vulnerabilities and Exposures. The MITRE Corporation. URL: https://cve.mitre.org.

Moller-Maersk, 2017. Interim Report Q3 2017. Technical Report 3. A.P Moller-Maersk A/S. Esplanaden 50, DK-1098 Copenhagen K. http://investor.maersk.com/static-files/1226cd7b-d1b4-4281-b42c-5f032b0e1595.

Moorhead, G., Griffin, R.W., 1998. Organizational behavior, managing people and organizations (5th ed.). Houghton Mifflin, Boston, MA, USA.

Nagaraju, V., Fiondella, L., Wandji, T., 2017. A survey of fault and attack tree modeling and analysis for cyber risk management, in: 2017 IEEE International Symposium on Technologies for Homeland Security (HST), IEEE. pp. 1–6.

Nahorney, B., 2019. SMB and the return of the worm. CISCO. Cisco Systems, Inc. Corporate Headquarters, 170 West Tasman Dr., San Jose, California 95134, U.S. https://blogs.cisco.com/security/smb-and-the-return-of-the-worm.

NCCIC, 2018. Security Tip (ST04-015) - Understanding Denial-of-Service Attacks. Cybersecurity and Infrastructure Security Agency. Rosslyn, Arlington, Virginia 20190, U.S. https://www.us-cert.gov/ncas/tips/ST04-015.

Norfund, 2020. Norfund has been exposed to a serious case of fraud. URL: https://www.norfund.no/norfund-has-been-exposed-to-a-serious-case-of-fraud/. [Online; accessed 18-May-2020].

Norman, D.A., 2002. The Design of Everyday Things. Basic Books, Inc.

Nowak, M., 2020. https://www.monterail.com/blog/why-ruby-on-rails-development-2020.

NSR, 2018. Mørketallsundersøkelsen 2018. Technical Report. Middelthuns gate 27, 0368 Oslo, Norway. https://www.nsr-org.no/getfile.php/1311303-1537281687/Bilder/M\OT1\orketallsunders\OT1\okelsen/M\OT1\orketallsunders\OT1\okelsen%202018%20low.pdf.

OWASP, 2020. OWASP Top Ten. OWASP. URL: https://owasp.org/www-project-top-ten/.

Pendse, S.G., 2012. Ethical hazards: A motive, means, and opportunity approach to curbing corporate unethical behavior. Journal of Business Ethics 107, 265–279.

Pols, P., 2017. The Unified Kill Chain. Technical Report. Cyber Security Academy (CSA). URL: https://www.csacademy.nl/images/scripties/2018/Paul_Pols_-_The_Unified_Kill_Chain_1.pdf. [Online; accessed 12-May-2020].

Psiaki, M.L., Humphreys, T.E., 2016. Gnss spoofing and detection. Proceedings of the IEEE 104, 1258–1270.

Ries, E., 2011. The lean startup : how constant innovation creates radically successful businesses. Portfolio Penguin.

Rogers, M., 2003. The role of criminal profiling in the computer forensics process. Computers & Security 22, 292–298.

Rogers, M.K., 2011. The psyche of cybercriminals: A psycho-social perspective, in: Cybercrimes: A multidisciplinary analysis. Springer, pp. 217–235.

Saini, V., Duan, Q., Paruchuri, V., 2008. Threat modeling using attack trees. Journal of Computing Sciences in Colleges 23, 124–131.

Schneier, B., 1999. Attack trees. Dr. Dobb's journal 24, 21–29.

Shang, W., Gong, T., Chen, C., Hou, J., Zeng, P., 2019. Information security risk assessment method for ship control system based on fuzzy sets and attack trees. Security and Communication Networks 2019.

Shaw, E.D., 2006. The role of behavioral research and profiling in malicious cyber insider investigations. Digital investigation 3, 20–31.

Shinder, D.L., Tittel, E., 2002. Chapter 3 - understanding the people on the scene, in: Scene of the Cybercrime. Syngress, Burlington, pp. 93 – 146. URL: http://www.sciencedirect.com/science/article/pii/B9781931836654500082, doi:https://doi.org/10.1016/B978-193183665-4/50008-2.

Shirey, R., 2007. Internet Security Glossary, Version 2. The IETF Trust. IETF Trust Internet Society, 11710 Plaza America Dr Suite 400, Reston, Virginia 20190, U.S. https://tools.ietf.org/html/rfc4949.

Simon, H.A., 1996. The Sciences of the Artificial (3rd Ed.). MIT Press, Cambridge, MA, USA.

Soegaard, M., 2018. The basics of user experience design. Interaction Design Foundation, ed https://www.interaction-design.org/ebook.

Sparks, S., Embleton, S., Zou, C.C., 2009. A chipset level network backdoor: bypassing host-based firewall & ids, in: Proceedings of the 4th International Symposium on Information, Computer, and Communications Security, pp. 125–134.

Strom, B.E., Applebaum, A., Miller, D.P., Nickels, K.C., Pennington, A.G., Thomas, C.B., 2018. Mitre att&ck: Design and philosophy. MITRE Product MP , 18–0944.

Strom, B.E., Battaglia, J.A., Kemmerer, M.S., Kupersanin, W., Miller, D.P., Wampler, C., Whitley, S.M., Wolf, R.D., 2017. Finding cyber threats with att&ck-based analytics. The MITRE Corporation, Tech. Rep. .

Symantec, 2017. Symantec 2019 Internet Security Threat Report. Technical Report. Symantec. https://docs.broadcom.com/doc/istr-22-2017-en.

Symantec, 2019. Symantec 2019 Internet Security Threat Report. Technical Report. Symantec. https://www-west.symantec.com/content/dam/symantec/docs/reports/istr-24-2019-en.pdf.

Tam, K., Jones, K., 2019. Macra: A model-based framework for maritime cyber-risk assessment. WMU Journal of Maritime Affairs 18, 129–163.

Tan, W.s., Liu, D., Bishu, R., 2009. Web evaluation: Heuristic evaluation vs. user testing. International Journal of Industrial Ergonomics 39, 621–627.

Van Ruitenbeek, E., Keefe, K., Sanders, W.H., Muehrcke, C., 2010. 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 (DSN 2010), pp. 17–18.

Warikoo, A., 2014. Proposed methodology for cyber criminal profiling. Information Security Journal: A Global Perspective 23, 172–178.

Wehinger, F., 2011. The dark net: Self-regulation dynamics of illegal online markets for identities and related services, in: 2011 European Intelligence and Security Informatics Conference, IEEE. pp. 209–213.

Wichmann, F.A., Sharpe, L.T., Gegenfurtner, K.R., 2002. The contributions of color to recognition memory for natural scenes. Journal of Experimental Psychology: Learning, Memory, and Cognition 28, 509.

Wikipedia, 2019a. https://twitter.com/Wikipedia/status/1170133355901251585?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed&ref_url=https%3A%2F%2Fwww.dw.com%2Fen%2Fmalicious-attack-takes-wikipedia-offline-in-germany%2Fa-50335521.

Wikipedia, 2019b. Arbitrary Code Execution. Wikipedia. https://en.wikipedia.org/wiki/Arbitrary_code_execution.

Wikipedia, 2019c. Convention over configuration. Wikipedia. https://en.wikipedia.org/wiki/Convention_over_configuration.

Wortman, P.A., Chandy, J.A., 2020. Smart: security model adversarial risk-based tool for systems security design evaluation. Journal of Cybersecurity 6, tyaa003.

Appendix A

Breaking the cyber kill chain by modelling resource costs

Breaking the cyber kill chain by modelling resource costs

Kristian Haga[1][0000−0001−9688−7029], Per Håkon Meland[1,2][0000−0002−5509−0184] and Guttorm Sindre[1][0000−0001−5739−8265]

[1] Norwegian University of Science and Technology, Norway
{kristian.haga,per.hakon.meland,guttorm.sindre}@ntnu.no
https://www.ntnu.no/
[2] SINTEF Digital, Norway
per.h.meland@sintef.no
https://www.sintef.no/

Abstract. To combat cybercrime, a clearer understanding of the attacks and the offenders is necessary. When there is little available data about attack incidents, which is usually the case for new technology, one can make estimations about the necessary investments an offender would need to compromise the system. The next step would be to implement measures that increase these costs to a level that makes the attack unattractive. Our research method follows the principles of *design science*, where cycles of research activities are used to create artefacts intended to solve real-world problems. Our artefacts are an approach for creating a *resource costs model* (RCM) and an accompanying modelling tool implemented as a web application. These are used to find the required attacker resources at each stage of the cyber kill chain. End user feedback show that structured visualisation of the required resources would raise the awareness of the cyberthreat. This approach has its strength and provides best accuracy with specific attacks, but is more limited when there are many possible attack vectors of different types.

Keywords: cyber kill chain, costs, resources, profiling, attack tree

1 Introduction

As our use of technology in almost every aspect of life steadily increases, so does our exposure to cybercrime. To combat this growing form of criminality, a clearer understanding of the costs, benefits and attractiveness of cyberattacks is necessary [16]. This is in accordance with *Routine Active Theory* [5], extended to include cybercrime [6, 8], which states that crime will occur when all of the following four conditions are met: There exist an *1) accessible and attractive target*, *2) the absence of a capable guardian* and the presence of *3) a motivated offender* with *4) the resources required to commit the crime.* For the latter case, it is not just a question of technical skills, but also a requirement that the offender is able to invest in software development and hardware acquisition, as well as the time it takes to plan, prepare and perform the attack. Alternatively, the offender could bribe an insider or hire someone else to do it through cybercrime-as-a-service [19] being offered by third parties.

We hypothesize that during threat analysis, it is possible to reduce the complexity of the resource requirement to a monetary concern, complemented by a limited set of attacker characteristics. This will allow us to identify the potential offenders and come up with technical and non-technical mitigations that will significantly increase the attacker costs.

The contribution of this paper is a modelling approach that maps resource costs to each stage of a cyberattack, and derives the total costs of the attack. We have utilized principles from Schneier's *attack trees* [30] and the Lockheed Martin's *cyber kill chain* [11], both already widely known in the security community, to structure this approach. A dedicated prototype tool has been developed to simplify and visualise this process, and we have completed the first rounds of iterative evaluation among experts. This tool is able to interactive show calculations and extract potential offenders based on a built-in library from available cybercriminal profile literature. Our goal is to improve the accuracy of threat analysis, and especially increase the understanding and awareness of cyberthreats among sectorial domain stakeholders.

This paper is structured as follows. Section 2 gives an overview of background knowledge and literature, and Section 3 explains our method. Results are given in Section 4, which are discussed in the light of evaluations in Section 5. Finally, Section 6 concludes the paper.

2 Background

2.1 The cyber kill chain

Already in 1998, Meadows [21] presented a way of dividing attacks into different stages or phases to make visual representation easier. The next stage would not commence before the previous one had completed, and she used different colours to represent the assumed difficulty of each stage. The stages were not predetermined, but varied according to the nature of the attack. Later on, McQueen et al. [20] defined a set of five fixed stages, *reconnaissance, breach, penetrate, escalation* and *damage*, which were then modelled as a compromise graph in order to find the weakest link(s) in the attack path based on expected time-to-compromise. Hutchins et al. [12] describe different phase-based models from military usage (countering terrorist attacks) and the information security field (between 2008-2010), and present their own version nicked the *intrusion kill chain.* This model was later on renamed and branded as the *cyber kill chain* [11] by Lockheed Martin, and has proven to be widely popular among defenders of IT and enterprise networks [1]. The seven stages of the cyber kill chain are:

1. **Reconnaissance** - Research, identification and selection of target.
2. **Weaponization** - Coupling a malware (e.g. remote access trojan) with an exploit into a deliverable payload, e.g. a media file.

3. **Delivery** - Transmission of the weapon to the targeted environment, e.g. an email attachment or USB-drive.
4. **Exploitation** - Triggers malicious code. Ranges from vulnerabilities or auto-executing features in host's operating system to users triggering execution.
5. **Installation** - Installation of the malware on the victim system, allowing the adversary to maintain presence inside the environment.
6. **Command and Control (C2)** - Establishes a channel for the adversary to access the target environment.
7. **Actions on Objectives** - Complete attack objectives, such as data extraction, break integrity or make system unavailable. Alternatively, establish a hop point to compromise additional systems.

As shown by Pols [25], there are many variants of the kill chain found in the literature. Some with different stage types and others with up to eighteen different stages. We chose to focus our work on the original seven stage cyber kill chain due to its popularity.

2.2 Attack tree cost modelling

Attack trees are acyclic graphs used to model threats from the viewpoint of the perpetrator. Schneier's original attack tree paper [30] showed how different costs could be assigned to alternative leaf nodes and how these propagated to define the cheapest way of attack. A fundamental paradigm for this kind of modelling is the assumption of a *rational attacker* [3], meaning that *1) there will be no attack if the attack is unprofitable* and *2) the attacker chooses the most profitable way of attacking.*

There have also been several approaches where costs are used in combination with other attributes. For instance, Buldas et al. [3] include costs, gains, penalties and associated probability values. Further examples of different attributes and references to papers that utilize costs in attack trees is given by Bagnato et al. [2]. Having more attributes enables additional ways of analysing attack trees, for instance Kumar et al. [17] show how to find the minimum time to complete an attack given a specific budget. Jensen et al. [13] present an approach where cost is a function of time instead of a constant cost per atomic attack attempt. Still, the major challenge of assigning accurate attribute values to attack tree nodes is difficult to overcome as attacker-specific information tends to be based on a best guess [29].

A comprehensive overview of more than thirty attack and defence modelling approaches based on directed acyclic graphs can be found in a survey paper by Kordy et al. [15]. A more recent survey focusing on fault and attack trees has been published by Nagaraju et al. [22].

2.3 Cybercriminal profiling

Shinder and Tittel [31] define a *profile* to be a set of characteristics likely to be shared by criminals who commit a certain type of crime. The use of profiles during criminal investigations can be traced several hundred years back in time, and though this is not an exact science, Nykodym et al. [23] argue that the track record legitimates the concept. However, they also argue that attackers have more advantages in a cyber setting as they do not have to be physically present at the crime scene.

The two main methods for profiling are known as *inductive* and *deductive* [34]. In the former, a profile database is developed based on information from already committed crime, and offender characteristics are correlated with types of crime. In the latter, forensics evidence is gathered from the crime scene and used to deduce the characteristics of the offender. Most of the established literature comes from the digital forensics field and relates to deductive profiling. We have been mostly interested in inductive profiling as a tool to identify potential offenders before any crime is actually committed. Furthermore, it is well established that likely offenders have a *motive*, *means* and *opportunity* (MMO) [33, 24] before committing any crime. As attacker costs belongs to the *means* characteristic, the literature becomes more limited. Warikoo et al. [34] have *capability factor* as one of their six profile identification metrics, where available resources for e.g. purchasing malware belongs. Preuß et al. [26] created a small set of profiles based on twelve cybercrime cases between 1998 and 2004. Due to the limited sample size, they could not create a structured set of attributes for these, but found that the principle of *minimum costs and maximum results* were present in all. Casey [4] presents a threat agent library of archetypal cybercriminal agents where *resources* is one of the eight attributes defining them. Casey's work is used to define *Attack Resource Level* in the cyberthreat exchange format *STIX* [14].

3 Method

Our research method follows the principles of *design science*, supporting a pragmatic research paradigm where artefacts are created to solve real-world problems by cycling through research activities related to *relevance*, *design* and *rigor* [32, 9]. The problem we try to address is the challenge of quantifying cyberrisks when there is little reliable historical data about attacks. Our artefacts are 1) an approach for creating a *resource costs model* (RCM), that is used to find the required attacker investments at each stage of the cyber kill chain and 2) an accompanying modelling tool implemented as a web application.

As a part of the relevance cycle, we initially worked with opportunities and problems related to cybersecurity for maritime shipping. We analysed typical vulnerabilities and threats towards eNavigation systems, and made cost estimations for attacking the various underlying technology modules.

During the rigor cycle, past knowledge, as presented in Section 2, was examined and we chose to build on practices that already had a significant uptake among practitioners.

Most central to design science research is the design cycle, consisting of artefact construction, evaluation and refinements based on feedback. Initially, we applied “pen-and-paper” variants of the RCM and validated the expressiveness by constructing models of known cyberattacks towards maritime systems. The second iteration produced a *minimum viable product* (MVP) of the tool. Ries [27] defines a MVP as the version of a new product which allows developers to collect the maximum amount of validated learning about customers with the least effort. Our MVP consisted of an info page tutorial and functionality for building basic resource costs models for each attack phase. For the evaluation we recruited eight security professionals who modelled a specific use case. These were observed during modelling and debriefed afterwards. The third iteration added the cybercriminal profiling feature, improved the user interface, as well as tweaking flawed features and functions.

4 Results

4.1 The resource costs model

In a *resource cost model* (RCM), each stage in the cyber kill chain represents the root node of a *resource tree*, depicted in Figure 1, which is similar in structure to an attack tree.

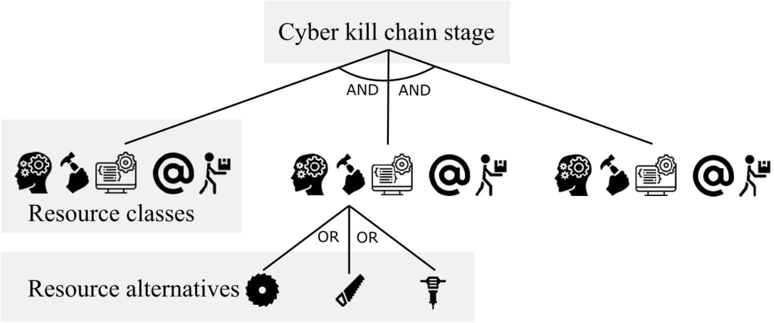


Fig. 1: A resource tree for a single cyber kill chain stage

The second level of the tree defines which resource types are required to complete the parent stage. At this level, all nodes have a conjunctive (AND) relationship since an attack would require all necessary resources. A resource can belong to five different classes:

Skill: Includes domain knowledge, malware development abilities or utilisation of cybercrime tools or guides.

Tangible: Necessary hardware components or other physical objects. This can range from advanced technology to soldering tools.

Logic: Commercially available software, data sets or cybercrime tools or services.

Logic-atomic: Necessary resources that can not be broken into smaller parts, e.g an IP-address, email address or a password.

Behavioral: Actions that must be conducted as a part of the attack, for instance bribing, sending out phishing emails or social engineering.

The third level in the tree, *resource alternatives*, are disjunctive (OR) leaf nodes that present ways to realize their parent resource class. Each resource alternative is associated with a cost interval and a confidence value. A confidence close to zero communicates that there is little evidence to support the stated cost interval. At the other end of the scale, a confidence of 1 means that there is exhaustive evidence to back the stated cost interval and that the price of the resource is not subject to great variation.

We can express the total cost interval of the attack T formally by stating that all resources R_j need to have a valid set V of resource alternatives. Let α represent the minimum estimated cost of the cheapest resource alternative and β represent maximum cost of the most expensive resource alternative. From this we can derive the following:

$$T = [(min\ cost = \sum_{\substack{\text{stage} \in \\ \text{kill chain}}} \sum_{i \in V} \alpha_i), (max\ cost = \sum_{\substack{\text{stage} \in \\ \text{kill chain}}} \sum_{i \in V} \beta_i)] \tag{1}$$

By letting ϕ be the average confidence of the n resource alternatives associated with a resource R_j and c_i is the confidence of a resource alternative i associated with R_j, we get the following associated confidence C of the total cost:

$$\phi_j = \frac{\sum_{i \in R_j} c_i}{n} \tag{2}$$

$$C = \prod_{\substack{\text{stage} \in \\ \text{kill chain}}} \prod_{R} \phi_j \tag{3}$$

In order to mitigate an attack, at least a one of the resources throughout the cyber kill chain must be made too expensive for the adversary. However, the adversary only needs a single resource alternative for each of the resources.

4.2 The IRCM tool

To validate the modelling approach, we have built an interactive installation of the model in the form of a web application called *Interactive Resource Cost Model* (IRCM) tool. This allows the users to model cyberattacks of their choosing, while concurrently deriving the total cost of the attack and probable cybercriminal profiles able to conduct it. An example screenshot from a single resource tree is shown in Figure 2, while a screenshot of the RCM for the complete cyber kill chain is included in Appendix A.

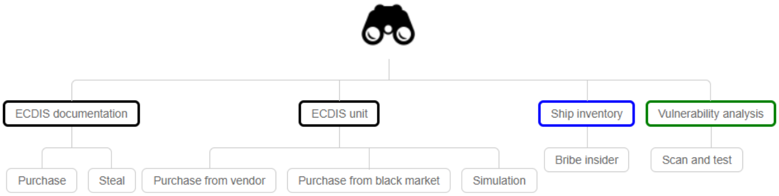
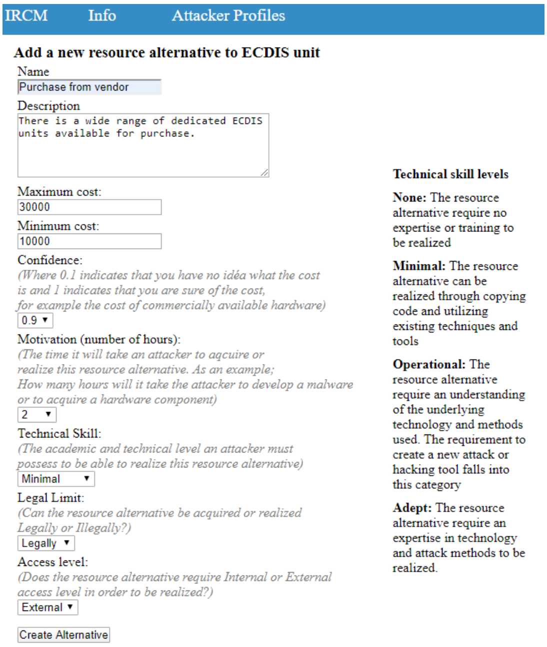


Fig. 2: A screenshot resource tree from the reconnaissance stage

These examples are taken from the maritime domain, where the *Electronic Chart Display and Information System* (ECDIS) is a central component for ship navigation. It displays the vessels position on a chart and integrates information from a number of sensors, such as radar, gyro, GNSS, echo sounder, weather measurements and the anti-collision systems. Malicious manipulation of this position could cause confusion on the ship bridge and potential course alteration could lead to collisions in congested waters [35]. The examples are loosely based on the demonstrated attack against an air-gapped ECDIS system by Lund et al. [18]. This attack was also structured according to the cyber kill chain, but in contrast to an external attack, it was conducted in cooperation with the Royal Norwegian Navy. Also, no information about resource costs were given, so here we have made our own estimations.

As can be seen in Figure 2, there are four resources defined for the reconnaissance stage. The first one, *ECDIS documentation*, is a tangible class, and the alternatives are to either *purchase* the documentation from the vendor legally, or *steal* it. The second resource is another tangible class, and represents an operational ECDIS unit that can be used to analyse its operating system, software and network traffic. It can be realized in different ways, by *purchasing a unit from vendor* or the *black market*, or running it as a software *simulation*. These alternatives vary in price, from relatively cheap software (where you pay according to sailing route) to more expensive hardware units in the range of \$10 000 - \$30 000. The third resource is of class logic-atomic, and represents information about the *ship inventory* used to determine which type and where the ECDIS units are installed. To simplify the model, only a single *bribe insider* alternative is used. The final resource is also of type skill, and represents required knowledge about *vulnerabilities* gained through *scanning and testing*.

Both resources and resource alternatives are created by using the tool input data forms. An example screenshot for the ECDIS resource alternative *purchased from vendor* is shown in Figure 3.



Add a new resource alternative to ECDIS unit

Name
Purchase from vendor

Description
There is a wide range of dedicated ECDIS units available for purchase.

Maximum cost:
30000

Minimum cost:
10000

Confidence:
(Where 0.1 indicates that you have no idéa what the cost is and 1 indicates that you are sure of the cost, for example the cost of commercially available hardware)
0.9 ▾

Motivation (number of hours):
(The time it will take an attacker to aqcuire or realize this resource alternative. As an example; How many hours will it take the attacker to develop a malware or to acquire a hardware component)
2 ▾

Technical Skill:
(The academic and technical level an attacker must possess to be able to realize this resource alternative)
Minimal ▾

Legal Limit:
(Can the resource alternative be acquired or realized Legally or Illegally?)
Legally ▾

Access level:
(Does the resource alternative require Internal or External access level in order to be realized?)
External ▾

Technical skill levels

None: The resource alternative require no expertise or training to be realized

Minimal: The resource alternative can be realized through copying code and utilizing existing techniques and tools

Operational: The resource alternative require an understanding of the underlying technology and methods used. The requirement to create a new attack or hacking tool falls into this category

Adept: The resource alternative require an expertise in technology and attack methods to be realized.

Fig. 3: A screenshot from the resource alternative window

The tool has a built-in database of cybercriminal profiles that the model inductively retrieves candidates from. This database is summarized in Appendix B and has been based on profile definitions we have found in the literature [34, 28, 4, 14]. We found out that mapping total attack cost with assumed *wealth* was not a very useful way of doing this. The wealthiest attacker is not always the most likely one, and attackers have more than one characterizing dimension. Therefore, the tool is able to exclude improbable attacker profiles from the database based on optional information that is assigned to the resources in the RCM. The exclusion rules are based on the following:

- Total minimal *cost* exceeds the financial capacities of the profile [*no cost, low, medium, high*].
- The accumulated time to require all resources exceed its *motivational* limit [*no time, low, medium, high*].
- Any resource alternative require a higher *technical skill* level than the profile possesses [*none, minimal, operational, adept*].
- Any resource that requires *moral limits* to be broken [*legally, illegally*].
- Any resource that require an *access level* the profile does not posses [*internal, external*].

The extended ECDIS attack example in Appendix A shows aggregated model information based on input contained in the individual resource tree for each attack stage. The cost interval has a broad range, mostly due to the choice of purchasing ECDIS hardware unit versus other cheaper alternatives in both the *reconnaissance* and *delivery* stages. Besides from these, the overall resource costs related to tangible and skill are relatively low. By analysing the model, we find that there are significant costs related to the *delivery* stage as the attacker would need physical presence at the ship and gain access to the bridge or bribe an insider. It is the air-gapping of the ECDIS that provides the main security measure by making delivery costly. When considering opening up for online software and chart updates, it is clear that additional secure measures will be needed to preserve an expensive attack vector. The confidence value is also very low, but would have been much higher if we had modelled the attack with a specific ECDIS unit in mind where costs are more certain. Also, a higher number of resources will automatically yield a lower confidence, which is natural since acquiring many resources increases uncertainty. The main benefit of the confidence is for attack comparison, which is not shown in these examples. Given the various exclusion rules that have been applied to the model, the most probable attacker profile in this case is *cyber warrior* (described in Appendix B).

5 Discussion

Hong and Kim [10] have pointed to the inherit challenge with graph-based attack models, namely the ability to scale. A purely tree-based model will generate large, bewildering attack trees for complex attacks. In turn, this creates a conflict between analysis and comprehensibility [7]. Hence, some sort of decomposition is needed. We chose to combine two modelling techniques to amplify their advantages and overcome some of their shortcomings. The cyber kill chain allows us to divide the attack into seven consecutive steps, and by stopping it in the early stages we don't have to embellish the later ones. The relatively small resource tree for each of the stages breaks down composite resource requirements into atomic ones, which can be more accurately estimated. This was the main takeaway from the first iteration of the design cycle. Secondly, we experienced that deriving a cost interval, rather than a single estimate, provides more confident information regarding the availability of an attack. A cheap, more available resource alternative set may provide a less stealthy attack than an expensive alternative. By determining both the minimum and maximum cost, we include both the risk willing and risk averse offenders. A large cost interval does not necessarily imply an inaccurate cost estimate, but rather that the evaluated attack can be carried out with a wide span of sophistication and possible impact on the target.

The second iteration involved expert end users who were observed using the MVP of the tool and debriefed afterwards. Seven out of these eight expressed that the main difficulty was to understand the difference between *resource* and *resource alternative* in the models. We were also able to observe that classifying resources was not straightforward, and the users spent some time navigating between the information page and the modelling interface to check definitions and the tutorial example. Both of these issues improved quickly with hands-on experience and by refining the info page. It was stated during the debrief that "especially interesting is the fact that making only a single resource unavailable, thus breaking the kill chain, will mitigate the entire attack" and all independently agreed that the structured visualisation of the required resources would raise the awareness of the cyberthreat. Some also expressed that many of the resources are impossible to make unavailable, which is true of course. In the MVP, we used *attack trees* as the tree structure term, and this caused some confusion since the RCM focus on resource required to perform the attack and not the attack actions, hence we changed this to *resource tree*.

The third iteration has had a focus on inducing criminal profiles from the models. As already mentioned, the wealthiest attacker is not always the most likely one, therefore we are using five identifying attributes as exclusion rules. A known limitation is that none of these say much about the *motive* of the offender, that is *why* would she commit the crime. This has been out of our scope, but could be extended by looking at the attack impact and attacker reward. Those considerations would have to be determined on a case-by-case basis, requiring additional knowledge dimensions. There is a general criticism towards the cyber kill chain that it focuses too much on the perimeter and malware attack vector [25], and we have seen supportive evidence of that too. Therefore, future improvements could be to include other sets of stages more suitable to describe attacks such as for instance related to social engineering, denial-of-service or code injection.

6 Conclusion

Through the iterative nature of design science we have made many improvements to the RCM modelling approach and the accompanying tool. However, we still consider this work to be in progress with many potential improvements related to usefulness and usability. We are also planning to extend the user testing and evaluation, particularly in the field of maritime cybersecurity, but also in other domains to ensure that the artefacts could have a wider usage than just the maritime context. Nevertheless, there is no silver bullet to threat modelling. We are trying to address the real-world problem of missing historical incident data, which is a particular concern for new technology. The RCM has its strength and provides best accuracy with specific attacks; when there are few resources and resource alternatives. Hence, we would not recommend this approach when you want to represent attacks with many possible attack vectors of different types. In such cases, several RCMs could be created and compared, but this quickly becomes a tedious task. As always, the analyst should choose the right tool for the job at hand.

Acknowledgment

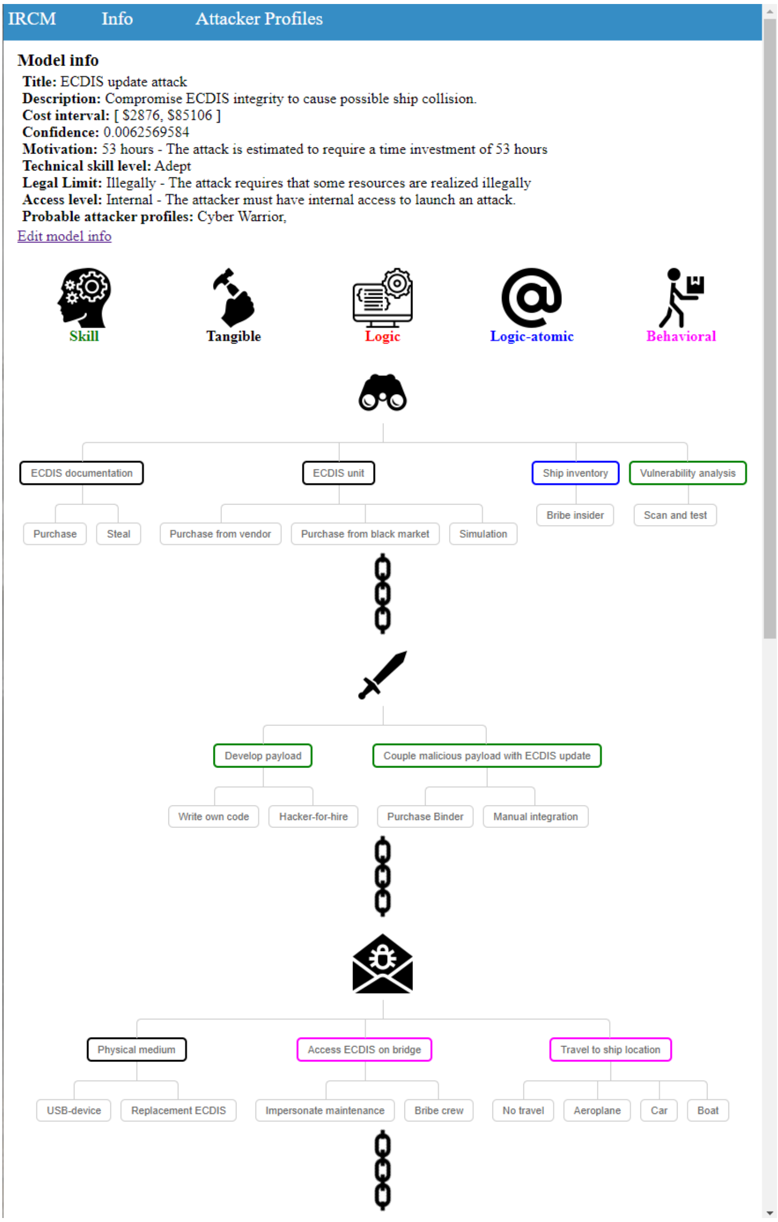
The research leading to these results has partially been performed by the Cyber Security in Merchant Shipping Service Evolution (CySiMS-SE) project, which received funding from the Research Council of Norway under Grant No. 295969.

References

1. Assante, M.J., Lee, R.M.: The industrial control system cyber kill chain. SANS Institute InfoSec Reading Room **1** (2015)
2. Bagnato, A., Kordy, B., Meland, P.H., Schweitzer, P.: Attribute decoration of attack–defense trees. International Journal of Secure Software Engineering (IJSSE) **3**(2), 1–35 (2012)
3. 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. pp. 235–248. Springer (2006)
4. Casey, T.: Threat agent library helps identify information security risks. Intel White Paper **2** (2007)
5. Cohoen, L.E., Felson, M.: Social change and crime rate trends: A routine activity approach. American Sociological Review **44**(4), 588–608 (1979)
6. Ekblom, P., Tiley, N.: Going equipped. The British Journal of Criminology **40**(3), 376–398 (2000)
7. Gadyatskaya, O., Trujillo-Rasua, R.: New directions in attack tree research: Catching up with industrial needs. In: International Workshop on Graphical Models for Security. pp. 115–126. Springer (2017)
8. Grabosky, P.N.: Virtual criminality: Old wine in new bottles? Social and Legal Studies **10**(2), 243–249 (2001)
9. Hevner, A., Chatterjee, S.: Design science research in information systems. In: Design research in information systems, pp. 9–22. Springer (2010)
10. Hong, J.B., Kim, D.S.: Performance analysis of scalable attack representation models. In: IFIP International Information Security Conference. pp. 330–343. Springer (2013)
11. Hutchins, E.M.: The cyber kill chain. Tech. rep., Lockheed Martin (2020), https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html, [Online; accessed 12-April-2020]
12. Hutchins, E.M., Cloppert, M.J., Amin, R.M.: Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. Leading Issues in Information Warfare & Security Research **1**(1), 80 (2011)
13. Jensen, P.G., Larsen, K., Legay, A., Poulsen, D.: Quantitative evaluation of attack defense trees using stochastic timed automata. In: International Workshop on Graphical Models for Security. pp. 75–90. HAL Id: hal-01640091 (2017)
14. Jordan, B., Piazza, R., Wounder, J.: Stix version 2.0. part 1: Stix core concepts. Tech. rep., OASIS Commettee Specifications 01 (2017), http://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part1-stix-core.html, [Online; accessed 13-April-2020]
15. Kordy, B., Piètre-Cambacédès, L., Schweitzer, P.: Dag-based attack and defense modeling: Don't miss the forest for the attack trees. Computer science review **13**, 1–38 (2014)
16. Kshetri, N.: The simple economics of cybercrimes. IEEE Security & Privacy **4**(1), 33–39 (2006)

17. Kumar, R., Ruijters, E., Stoelinga, M.: Quantitative attack tree analysis via priced timed automata. In: International Conference on Formal Modeling and Analysis of Timed Systems. pp. 156–171. Springer (2015)
18. Lund, M.S., Hareide, O.S., Jøsok, Ø.: An attack on an integrated navigation system. Necesse **3**(2), 149–163 (2018)
19. Manky, D.: Cybercrime as a service: a very modern business. Computer Fraud & Security **2013**(6), 9–13 (2013)
20. McQueen, M.A., Boyer, W.F., Flynn, M.A., Beitel, G.A.: Quantitative cyber risk reduction estimation methodology for a small scada control system. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06). vol. 9, pp. 226–226. IEEE (2006)
21. Meadows, C.: A representation of protocol attacks for risk assessment. In: Proceedings of the DIMACS Workshop on Network Threats. pp. 1–10 (1998)
22. Nagaraju, V., Fiondella, L., Wandji, T.: A survey of fault and attack tree modeling and analysis for cyber risk management. In: 2017 IEEE International Symposium on Technologies for Homeland Security (HST). pp. 1–6. IEEE (2017)
23. Nykodym, N., Taylor, R., Vilela, J.: Criminal profiling and insider cyber crime. Computer Law & Security Review **21**(5), 408–414 (2005)
24. Pendse, S.G.: Ethical hazards: A motive, means, and opportunity approach to curbing corporate unethical behavior. Journal of Business Ethics **107**(3), 265–279 (2012)
25. Pols, P.: The unified kill chain: Designing a unified kill chain for analyzing, comparing and defending against cyber attacks. Cyber Security Academy (2017)
26. Preuß, J., Furnell, S.M., Papadaki, M.: Considering the potential of criminal profiling to combat hacking. Journal in Computer Virology **3**(2), 135–141 (2007)
27. Ries, E.: The lean startup : how constant innovation creates radically successful businesses. Portfolio Penguin (2011)
28. Rogers, M.K.: The psyche of cybercriminals: A psycho-social perspective. In: Cybercrimes: A multidisciplinary analysis, pp. 217–235. Springer (2011)
29. Saini, V., Duan, Q., Paruchuri, V.: Threat modeling using attack trees. Journal of Computing Sciences in Colleges **23**(4), 124–131 (2008)
30. Schneier, B.: Attack trees. Dr. Dobb's journal **24**(12), 21–29 (1999)
31. Shinder, D.L., Tittel, E.: Chapter 3 - understanding the people on the scene. In: Scene of the Cybercrime, pp. 93 – 146. Syngress, Burlington (2002)
32. Simon, H.A.: The Sciences of the Artificial (3rd Ed.). MIT Press, Cambridge, MA, USA (1996)
33. Van Ruitenbeek, E., Keefe, K., Sanders, W.H., 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 (DSN 2010). pp. 17–18 (2010)
34. Warikoo, A.: Proposed methodology for cyber criminal profiling. Information Security Journal: A Global Perspective **23**(4-6), 172–178 (2014)
35. Wingrove, M.: Security flaws open ECDIS to cyber crime. Tech. rep., Riviera (2018), https://www.rivieramm.com/opinion/opinion/security-flaws-open-ecdis-to-cyber-crime-24334, [Online; accessed 20-April-2020]

A Tool screenshots



Model info

Title: ECDIS update attack
Description: Compromise ECDIS integrity to cause possible ship collision.
Cost interval: [$2876, $85106]
Confidence: 0.0062569584
Motivation: 53 hours - The attack is estimated to require a time investment of 53 hours
Technical skill level: Adept
Legal Limit: Illegally - The attack requires that some resources are realized illegally
Access level: Internal - The attacker must have internal access to launch an attack.
Probable attacker profiles: Cyber Warrior,

Edit model info

Fig. 4: A screenshot from the first three stages; *Reconnaissance*, *Weaponization* and *Delivery*.

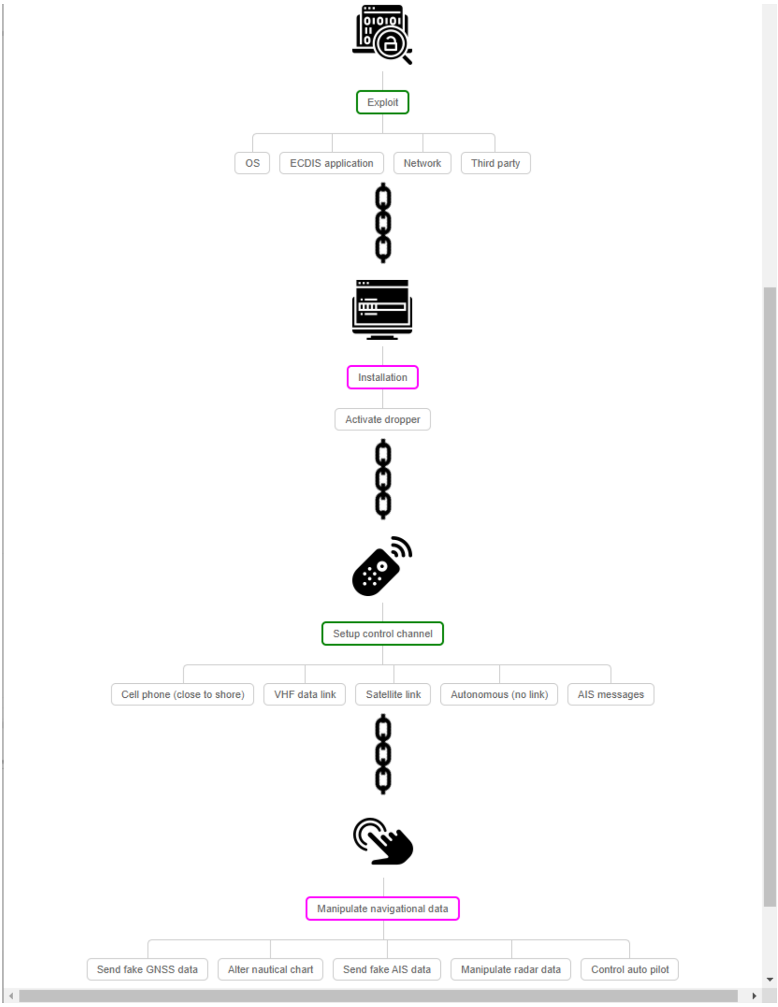


Fig. 5: A screenshot from the last four stages; *Exploitation*, *Installation*, *Command and Control* and *Actions on Objectives*

B Cybercriminal profiles

Script kiddie (SK) has a low level of motivation, thus *time consuming* attacks are not attractive to this profile. The technical skills are limited to *minimal* and the profile only accepts a *minimal* cost. Script kiddies will only utilize resources that can be realized *legally* and have *external* access.

Hacktivist (H) has a medium to high level of motivation anchored in the political cause they represent, thus they may conduct *time consuming*, targeted attacks. The technical skills of a hacktivist is limited to *minimal.* In order to fight for their cause, the hacktivist accepts *some* expenses. The hacktivist is willing to require resources *illegally* and have *external* access level.

Vandal (V) has a low to medium motivation and will only invest a *limited* amount of time in attention seeking attacks. The technical skills of the vandal is limited to *minimal* and the profile accepts a *low* cost. Vandals will only utilize resources that can be realized *legally* and have *external* access.

Petty criminal (PC) has a medium motivation level, willing to invest *some* time in attacks that bring financial gain. They possess *operational* technical skills and accepts a *medium* cost. The petty criminal is willing to require resources *illegally* and has *external* access level.

Mobster (M) has a medium to high level of motivation given that financial gain is possible, thus they may conduct *time consuming* attacks. The technical skills are *operational* and the profile accepts *costly* attacks. Mobsters won't second guess *illegal* resources and have *external* access level.

Cyberwarrior (CW) is a state-sponsored actor with a high motivation level, thus will conduct persistent, *highly time consuming* attacks. The cyberwarrior has *adept* technical skills for launching any attack. In addition, the cyberwarrior is *not limited* by any costs and disposes resources that may be required *illegally.* As an immediate result of the *adept* skill level, the cyberwarrior has *internal* access.

Terrorist (T) tends to be highly motivated and well funded, thus can conduct *time consuming* and *costly* cyberattacks to front beliefs. The technical skills are limited to *minimal.* The Terrorist is willing to require resources *illegally* and have *external* access level.

Internal - Hostile (IN-H) has a medium motivation level and may launch attacks that require *some* time. The profile knows the system well, which yields an *operational* technical skill. *Some* expenses are acceptable, limited to *legally* acquired resources. Internals have *internal* access level by default.

Internal - Non-hostile (IN-NH) launces cyberattacks by accident, thus *not* motivated at all to invest any time or money in a cyberattack and will only possess resources that can be *legally* realized. Given that accidental cyberattacks are possible yields an *operational* skill level and an *internal* access level.

Appendix B

RCM Pen and Paper Validation

RCM was validated in the preliminary work by modelling a GNSS spoofing attack on maritime vessels.

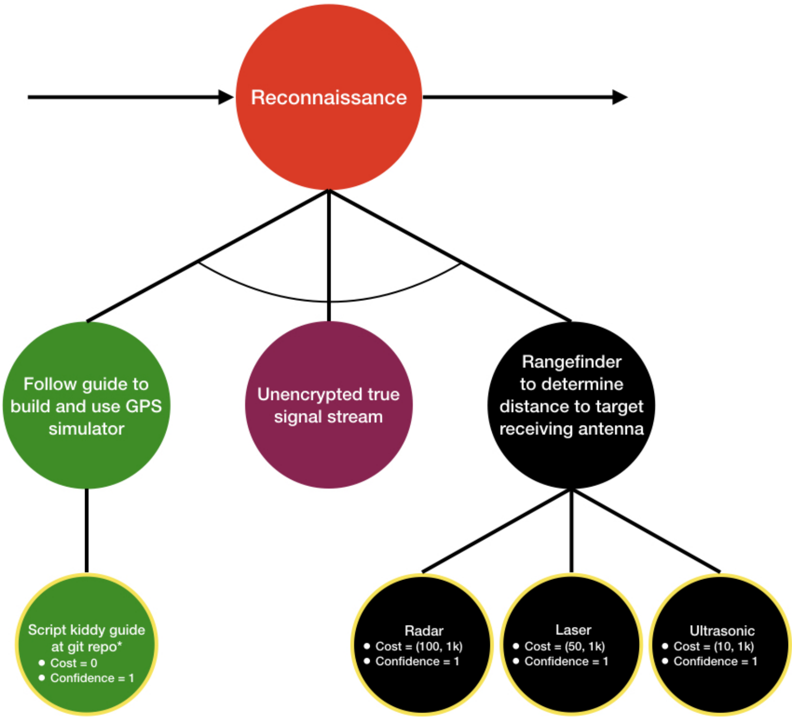

Reconnaissance
Follow guide to build and use GPS simulator
Unencrypted true signal stream
Rangefinder to determine distance to target receiving antenna
Script kiddy guide at git repo*
Cost = 0
Confidence = 1
Radar
Cost = (100, 1k)
Confidence = 1
Laser
Cost = (50, 1k)
Confidence = 1
Ultrasonic
Cost = (10, 1k)
Confidence = 1

Figure B.1: RCM Pen and Paper Reconnaissance stage

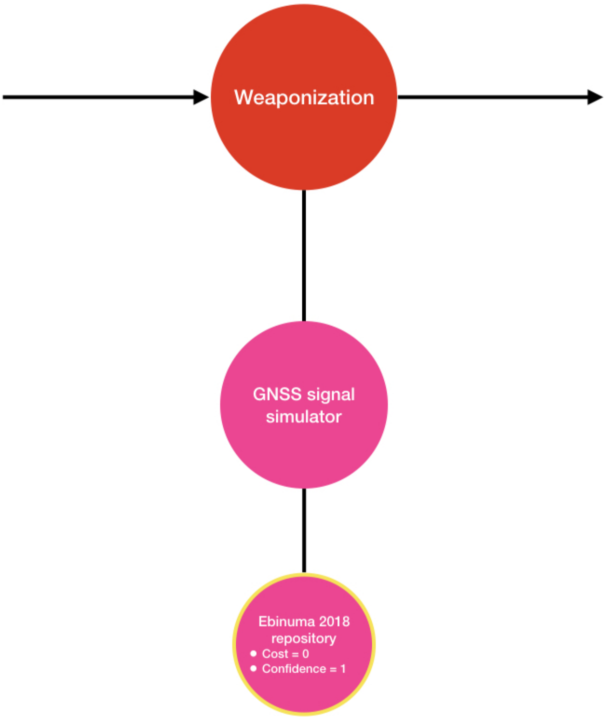


Figure B.2: RCM Pen and Paper Weaponization stage

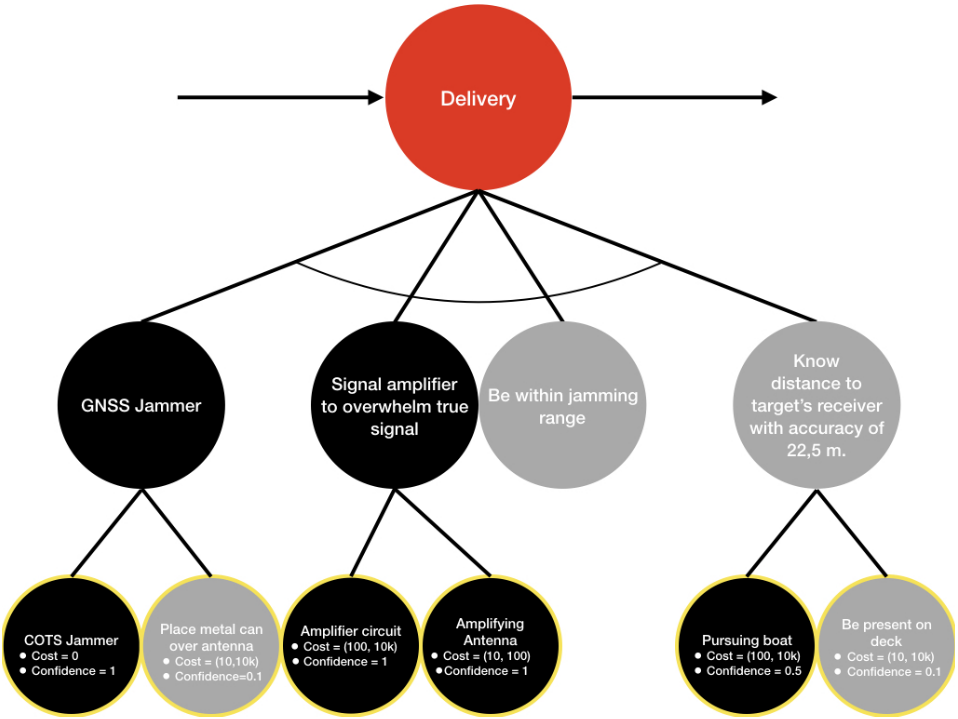


Figure B.3: RCM Pen and Paper Delivery stage

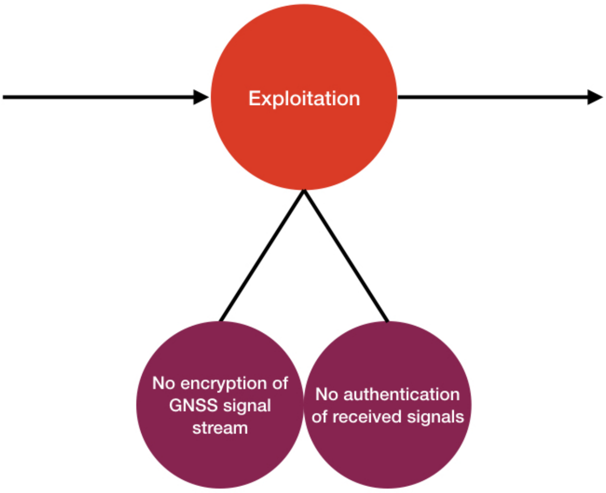

Exploitation
No encryption of GNSS signal stream
No authentication of received signals

Figure B.4: RCM Pen and Paper Exploitation stage

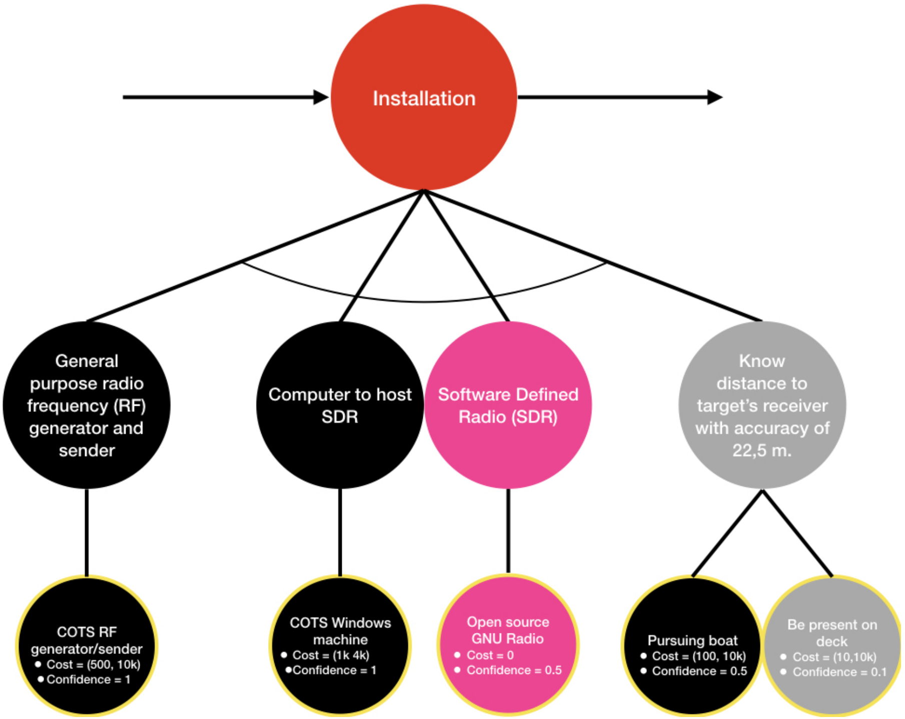


Figure B.5: RCM Pen and Paper Installation stage

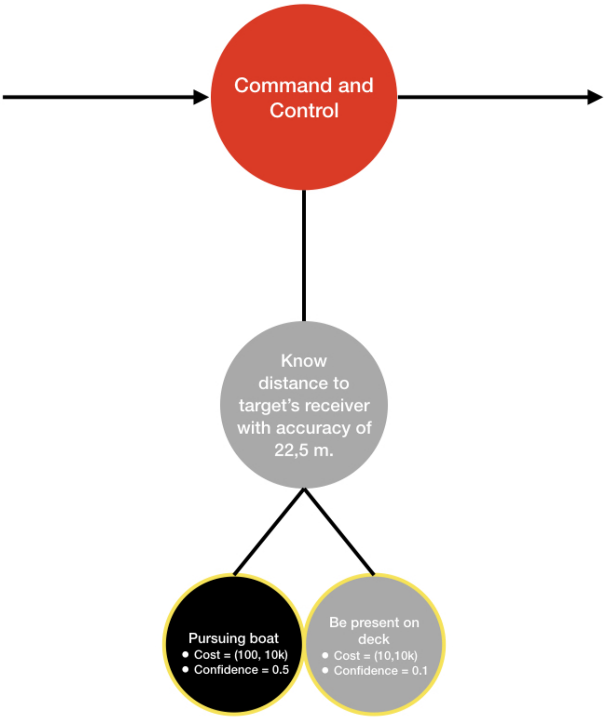


Figure B.6: RCM Pen and Paper Command and Control stage

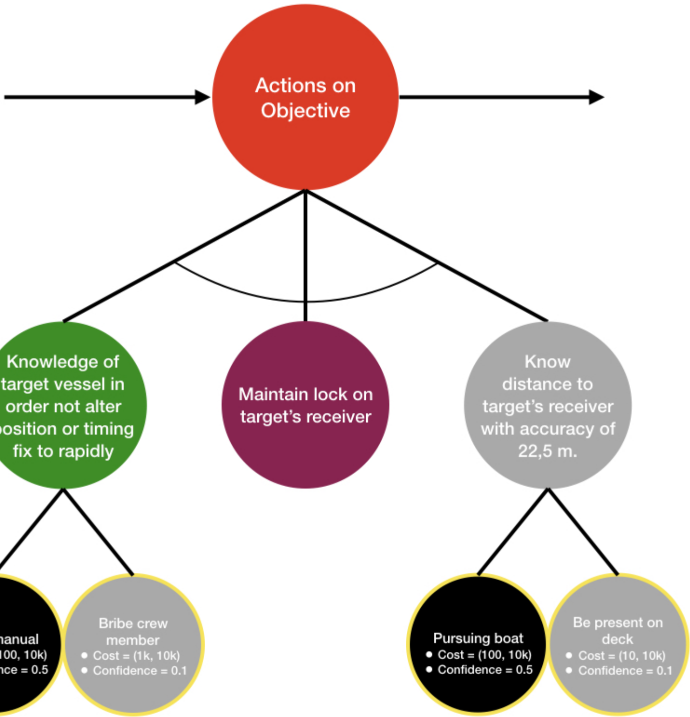


Figure B.7: RCM Pen and Paper Actions on Objective stage

Appendix C

User Test Cases

C.1 Cyberattack case - GNSS Spoofing

The case presented to the test subjects was a GNSS Spoofing attack on maritime vessels.

A GNSS spoofing attack broadcast false GNSS signals with the intent that the victim will misinterpret them as authentic signals. The literature presents three main attack vectors to perform GNSS spoofing by altering either position or time data. One method involve inducing the target receiver to lock onto a malicious signal stream. This is achieved by forcing the target to re-acquire its signal lock, e.g by jamming the signal, before overwhelming the true signal with a synthesized, malicious signal sent at a greater strength resulting in the target receiver locking on the malicious signal. A re-acquisition and signal lock can be performed undetected when the attacker knows the distance to the victim with an accuracy of 22.5 m. The second method, known as meaconing, records the true GNSS signals and replays the signal at overwhelming strength with a delay, thus altering the deduced location. Meaconing has the potential of spoofing any GNSS signal, even encrypted military signals. The third method, called nulling, involves the adversary generating two signal streams. One stream being the negative of the true signal, thus cancelling out the true signal. The other stream consisting of spoofed signals.

Below we present the required resources associated with each stage of the kill chain to conduct the attack.

Reconnaissance

All three attack methods rely on the malicious agent being within a certain range of the target's receiving antenna. Hence, the agent must have knowledge on the location of the GNSS receiving antenna of the victim.

Nulling and meaconing both require the attacker to have subtle knowledge on GNSS signal processing in order to synthesize the undetectable malicious signal stream. This information is publicly available.

Weaponization

All three attack methods require the ability to simulate a GNSS signal in order to synthesize a false signal stream. A GPS simulator proven to work for GPS spoofing attacks is publicly accessible on Github.

Delivery

The delivery stage of a GPS spoofing attack involves the target's receiving antenna to receive the false, malicious signal stream instead of the true signal stream. This could be achieved through a) a forced re-acquisition caused by jamming, b) the false signal overwhelming the true signal by being sent at a greater strength or c) by generating a signal stream canceling the true stream and introducing a false stream.

Exploitation

Civilian GNSS signals are not encrypted or authenticated, thus an adversary forcing the target receiver to read its signal stream could spoof the target GNSS. Attacks on military encrypted, authenticated systems can only read and replay signals. Such attacks exploit features in the radio signal in order to alter the deduced timing or position data.

Installation

The installation stage of GNSS spoofing consists of the target interpreting and processing the synthesized, false signal stream, perceiving it as the true stream. To synthesize a GNSS signal stream the attacker can use a Software Defined Radio (SDR) to convert the false, simulated GNSS stream into radio frequencies (RF). Then the attacker needs a general purpose RF generator and sender in order to send the false signal stream.

Command and Control

In order for the attacker to continuously control the target system, it must be able to maintain the connection between the false stream and target's receiving antenna. This require the attacker to know the position of the target with an accuracy of 22.5 m and to be within sending range of the chosen RF generator. This could be achieved by being present on deck or pursuing the target by boat or drone.

Action on Objective

In order to not reveal the undergoing attack the adversary should not attempt to alter the victim's position or timing fix too rapidly. Also, the adversary will have to take target specific parameters into account, e.g not set the false speed above the top speed of the target. A stealthy attack should induce a false stream that is within the margins of error of other navigation systems, e.g magnetic compass and wind measurements, on the target vessel, thus ensuring that these systems do not set of any alarms warning about the unnatural change in deduced location.

C.2 UT2.A Case

A GPS spoofing attack seeks to force the target navigation system to display a position set by the attacker instead of the true position of the target, meaning the actual location of the target. The consequences of a GPS spoofing attack on a ship is that the target is put off course, which may lead to the target ship going down and quotes like ”TS Sola, DU MÅ DREIA!!”.

The attack is accomplished by the attacker broadcasting a false GPS signal stream and the target interpreting the false stream as authentic signals. This can be achieved by first jamming the true GPS signal of the target, which forces the target system to re-acquire the GPS signal. During this re-acquisition the attacker broadcast the false signal stream at a greater strength than the true stream, resulting in the target connecting to the stronger, false stream.

When the target is connected to the false stream, the attacker must ensure a persistent connection and that the target system doesn't detect any errors in the false signal. Such an error could be that the false signal stream indicates that the target ship is moving at a speed greater than known top speed.

C.2.1 Attacking KNM Ingstad Helge

After a dreadful year of service in the Norwegian Navy, before moving to the Utopia known as Gløshaugen to pursue a career in cybersecurity, you seek your revenge and a big payday to make up for the insufficient student loan from Lånekassen. To achieve this you team up with the captain of ”TS Månen” to set up a bullet proof insurance fraud. By causing the navy vessel ”KNM Ingstad Helge” to crash into ”TS Månen”, you will get your revenge and the pay-out from the insurance of TS Månen will buy you endless beer in the Bodega.

Using a kill chain you derive the required resources to carry out a GPS Spoofing attack.

Reconnaissance

You must know the true position of the GPS antenna at KNM Ingstad Helge with an accuracy of 22.5 meters in order to generate a valid false signal - a signal not being detected as false by the navigation system on KNM Ingstad Helge. To synthesize a valid false signal stream you must also have subtle knowledge on GPS signal processing. You can learn about GPS signal processing in the course TDT1337 or researching public available information on GPS signal processing online.

Weaponization

For the Weaponization stage you will need to synthesize a false GPS signal stream. To do the actual synthesizing of the false signal you will need a GPS Signal Simulator. The simulator is a software that converts a coordinate position into a GPS signal. There are several known working GPS Signal Simulators available at Github.

Delivery

The Delivery stage of the attack consist of forcing the target to receive the false signal stream. To achieve this you must force the target to do a signal re-acquisition. This will cause the target system to switch form the true signal stream to the false stream you control. An option to force the target navigation system on KNM Ingstad Helge to do a re-acquisition, is to use a GPS Jammer to jam the GPS navigation system. It is possible to buy weak GPS jammers at Biltema for $100-$500 or it is possible to buy military-grade, strong jammers on the Darkweb for $1000-$3000.

After jamming the target GPS system, when the system is re-acquiring a GPS signal, you must ensure that the system reconnects to the false stream. This can be done by sending the false signal at a greater strength than the true signal. For this you will need an amplifying antenna. Such antennas are available at Amazon for $10-$100.

Exploitation

Unfortunately you lost your encryption key to the Norwegian Navy GPS signals. As a result, to be able to carry out the attack, you must rely on KNM Ingstad Helge using an unencrypted GPS signal. This atomic fact can be assured by bribing a navigation officer to turn off the encryption. Luckily, Norwegian officers are loyal and a bribe might come at a high cost, but you are not sure just how high. An option could be to blackmail an officer you know to have stolen gear from the navy.

Installation

The Installation stage involves to send the false signal and the target to process this. To send the signal you will need a Software Defined Radio (SDR), a general purpose Radio Frequency Generator (RFG) and a Radio Signal Sender (RSS). The SDR converts the synthesized false GPS signal stream to radio frequencies, before the RFG generates the actual signal to be sent by the RSS to KNM Ingstad Helge.

GNU offers a free SDR for download online, while the latter two hardware components are publicly available online. Both are within the prize range $1000-$5000. In order to connect the three components needed to send the signal you will need a laptop running Windows.

Command and Control - C2

To maintain control over the target system you will need to continuously send a valid false GPS signal stream. This requires you to continuously know the position of KNM Ingstad Helge with an accuracy of 22.5 meters and to be within sending range of the RSS. To ensure this you can sneak onboard the ship, but an easier option might be to pursuit the target ship either by drone or boat.

Actions on Objective - AO

As your intentions are to misguide KNM Ingstad Helge for a significant period of time in order to crash into TS Månen, it is crucial that the malicious actions you take during the attack goes undetected. To ensure that neither the navigation system nor any crew members onboard KNM Ingstad Helge detect the ongoing attack, you must not alter the false position to rapidly. This may lead to unnatural behavior in the navigation system, for example the calculated speed being more than the top speed of KNM Ingstad Helge or that an interpreted change in direction is beyond the capacity of the ship.

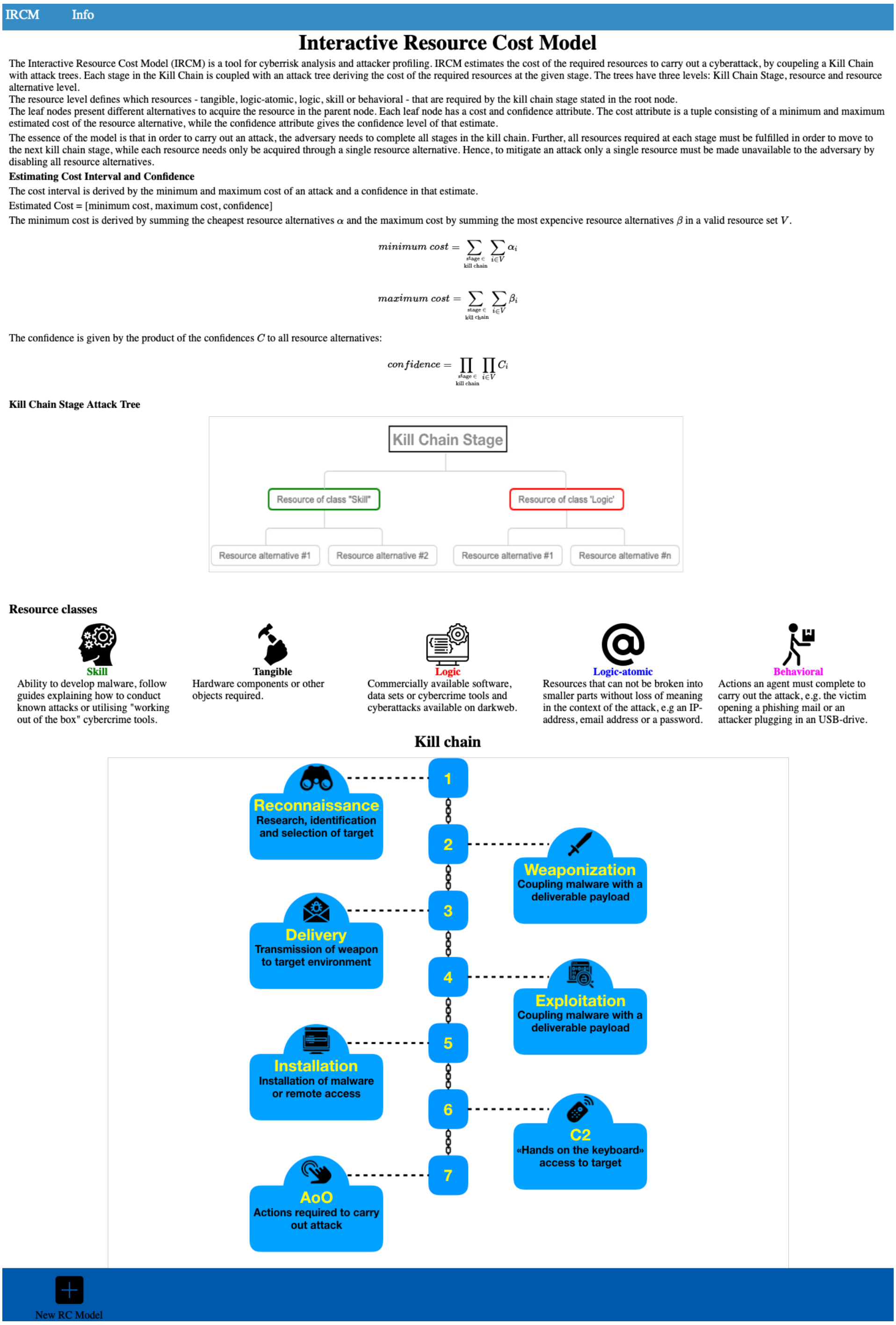
To avoid such errors you need to get your hands on the KNM Ingstad Helge user manual. Where can you possibly find this?

Appendix D

IRCM screenshots

D.1 IRCM MVP

Below are screenshots of the IRCM MVP used in User Test 1, see 7.1.



Interactive Resource Cost Model

The Interactive Resource Cost Model (IRCM) is a tool for cyberrisk analysis and attacker profiling. IRCM estimates the cost of the required resources to carry out a cyberattack, by coupeling a Kill Chain with attack trees. Each stage in the Kill Chain is coupled with an attack tree deriving the cost of the required resources at the given stage. The trees have three levels: Kill Chain Stage, resource and resource alternative level.

The resource level defines which resources - tangible, logic-atomic, logic, skill or behavioral - that are required by the kill chain stage stated in the root node.

The leaf nodes present different alternatives to acquire the resource in the parent node. Each leaf node has a cost and confidence attribute. The cost attribute is a tuple consisting of a minimum and maximum estimated cost of the resource alternative, while the confidence attribute gives the confidence level of that estimate.

The essence of the model is that in order to carry out an attack, the adversary needs to complete all stages in the kill chain. Further, all resources required at each stage must be fulfilled in order to move to the next kill chain stage, while each resource needs only be acquired through a single resource alternative. Hence, to mitigate an attack only a single resource must be made unavailable to the adversary by disabling all resource alternatives.

Estimating Cost Interval and Confidence

The cost interval is derived by the minimum and maximum cost of an attack and a confidence in that estimate.

Estimated Cost = [minimum cost, maximum cost, confidence]

The minimum cost is derived by summing the cheapest resource alternatives α and the maximum cost by summing the most expencive resource alternatives β in a valid resource set V.

$$minimum\ cost = \sum_{\substack{\text{stage} \in \\ \text{kill chain}}} \sum_{i \in V} \alpha_i$$

$$maximum\ cost = \sum_{\substack{\text{stage} \in \\ \text{kill chain}}} \sum_{i \in V} \beta_i$$

The confidence is given by the product of the confidences C to all resource alternatives:

$$confidence = \prod_{\substack{\text{stage} \in \\ \text{kill chain}}} \prod_{i \in V} C_i$$

Kill Chain Stage Attack Tree

Resource classes

Skill
Ability to develop malware, follow guides explaining how to conduct known attacks or utilising "working out of the box" cybercrime tools.

Tangible
Hardware components or other objects required.

Logic
Commercially available software, data sets or cybercrime tools and cyberattacks available on darkweb.

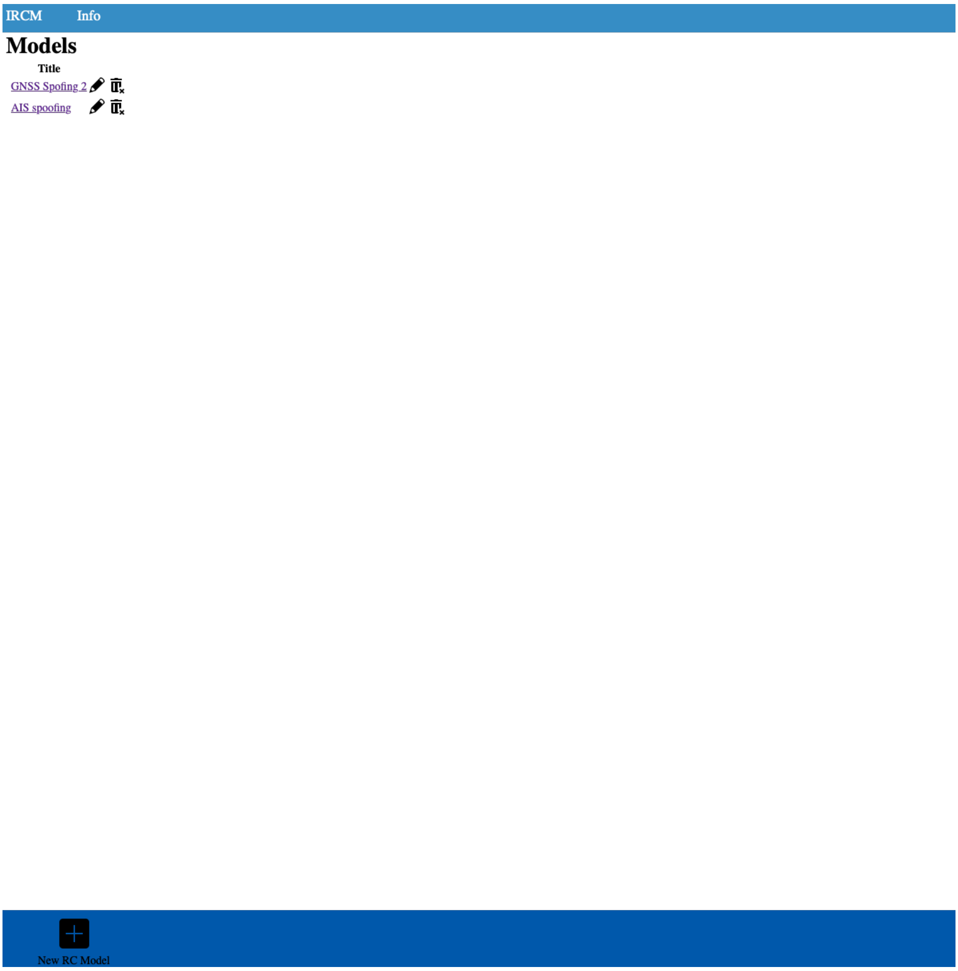
Logic-atomic
Resources that can not be broken into smaller parts without loss of meaning in the context of the attack, e.g an IP-address, email address or a password.

Behavioral
Actions an agent must complete to carry out the attack, e.g. the victim opening a phishing mail or an attacker plugging in an USB-drive.

Kill chain

1
Reconnaissance
Research, identification and selection of target
2
Weaponization
Coupling malware with a deliverable payload
3
Delivery
Transmission of weapon to target environment
4
Exploitation
Coupling malware with a deliverable payload
5
Installation
Installation of malware or remote access
6
C2
«Hands on the keyboard» access to target
7
AoO
Actions required to carry out attack
New RC Model

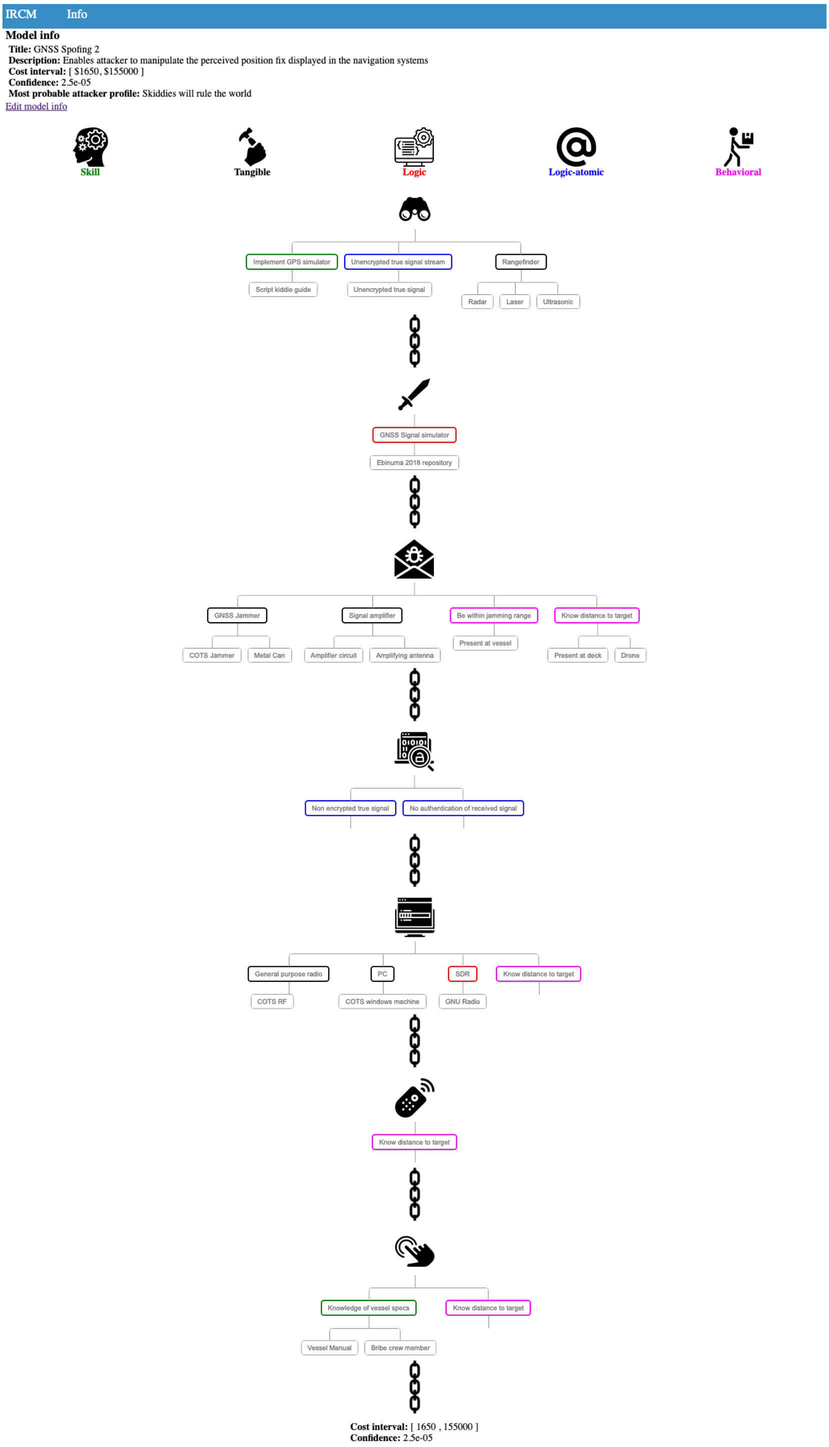
Figure D.1: Info-page



Models

Title		
GNSS Spofing 2	✎	🗑
AIS spoofing	✎	🗑

Figure D.2: Index page listing RC Models made by the user



IRCM Info

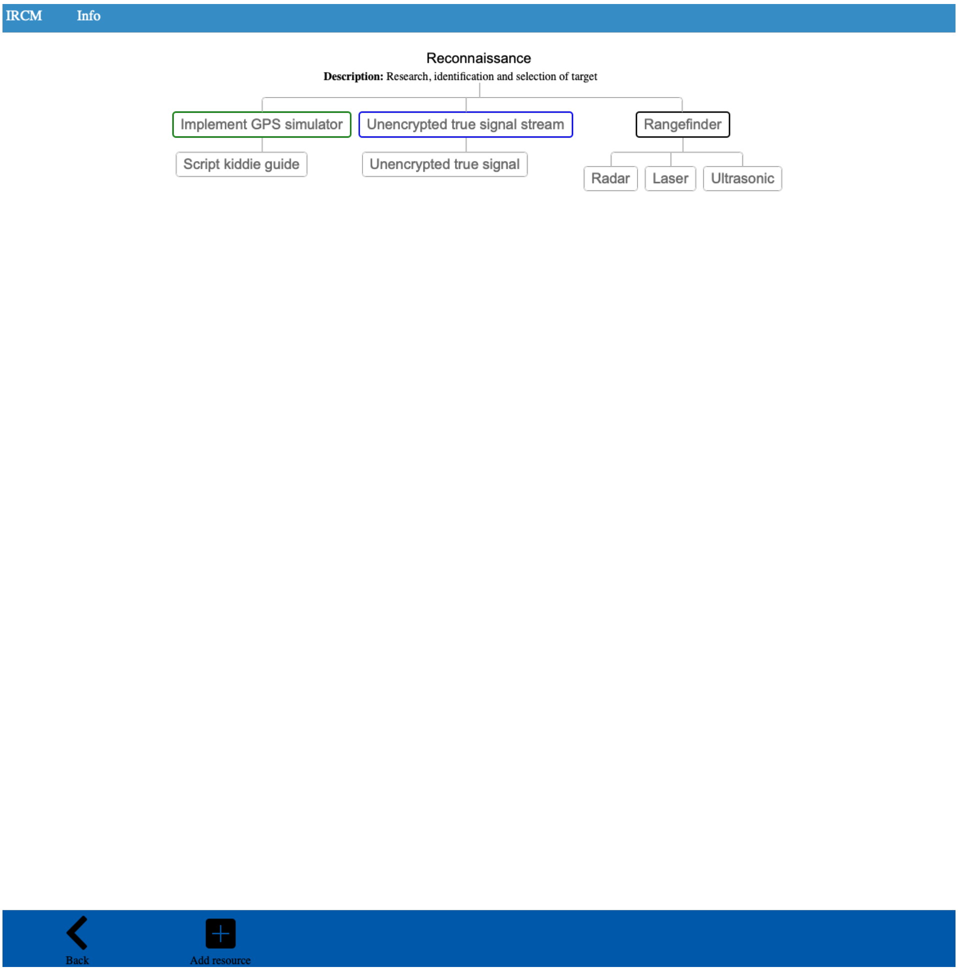
Model info

Title: GNSS Spofing 2
Description: Enables attacker to manipulate the perceived position fix displayed in the navigation systems
Cost interval: [$1650, $155000]
Confidence: 2.5e-05
Most probable attacker profile: Skiddies will rule the world

Edit model info

Cost interval: [1650 , 155000]
Confidence: 2.5e-05

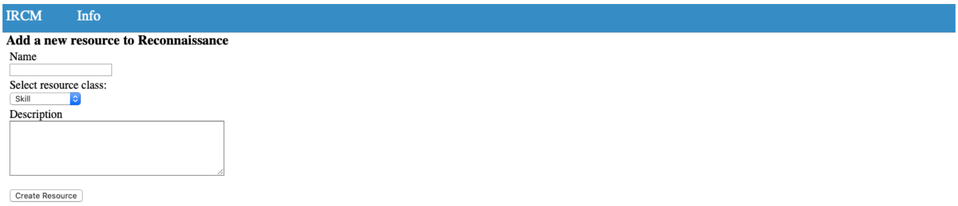
Figure D.3: Populated RC Model modeling a GNSS attack



Reconnaissance

Description: Research, identification and selection of target

Figure D.4: Page displaying the resource tree associated with the Reconnaissance stage - Similar pages were available for all stages



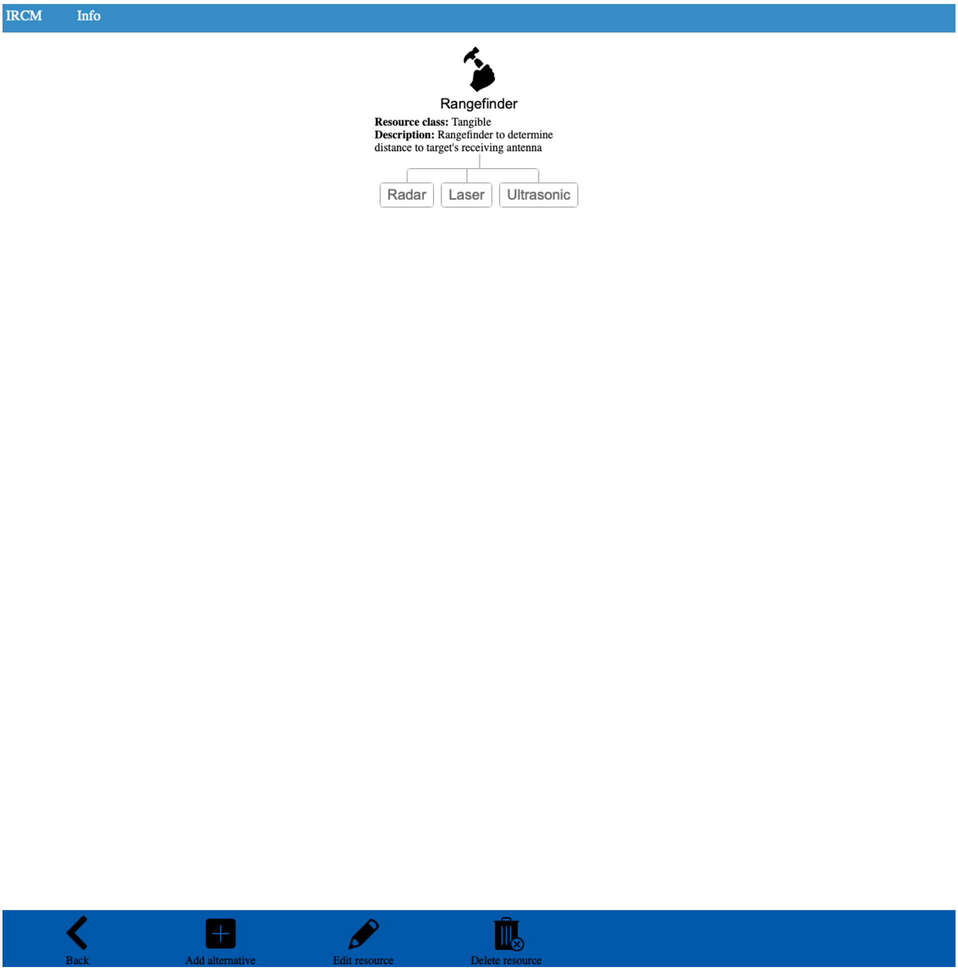
Add a new resource to Reconnaissance

Name

Select resource class:
Skill

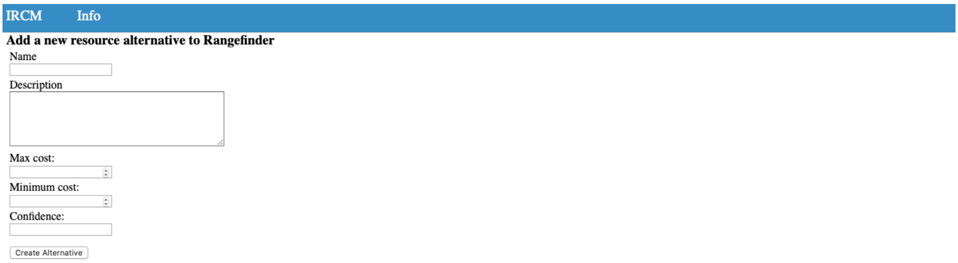
Description

Figure D.5: Page for adding a resource. In order to add a resource the user first had to select the stage to add the resource by viewing page shown in D.4 before clicking ”Add resource”. The same page is used for editing a resource



Resource class: Tangible
Description: Rangefinder to determine distance to target's receiving antenna

Figure D.6: Page displaying an already added resource



Add a new resource alternative to Rangefinder

Name

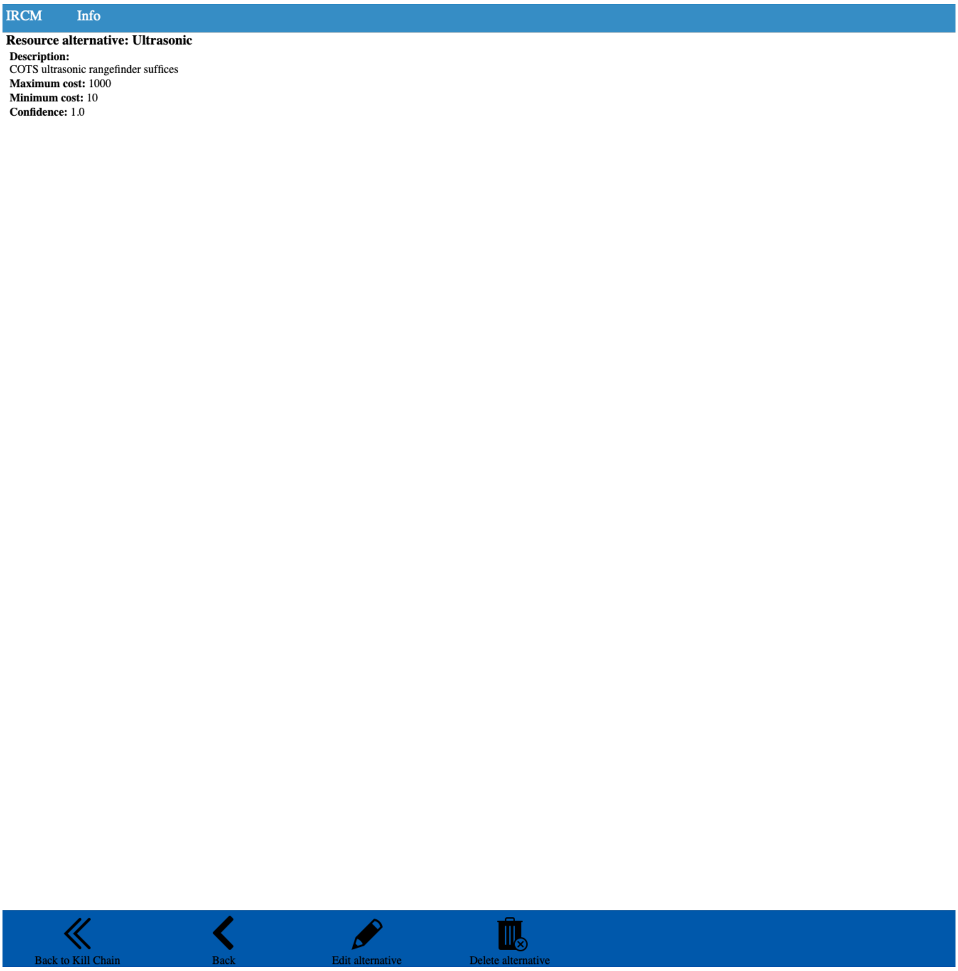
Description

Max cost:

Minimum cost:

Confidence:

Figure D.7: Page enabling the user to add a resource alternative. The user first selected the resource before being able to add a resource alternative by clicking ”Add alternative” shown is D.6. The same page is used for editing a resource alternative.



Resource alternative: Ultrasonic

Description:
COTS ultrasonic rangefinder suffices

Maximum cost: 1000

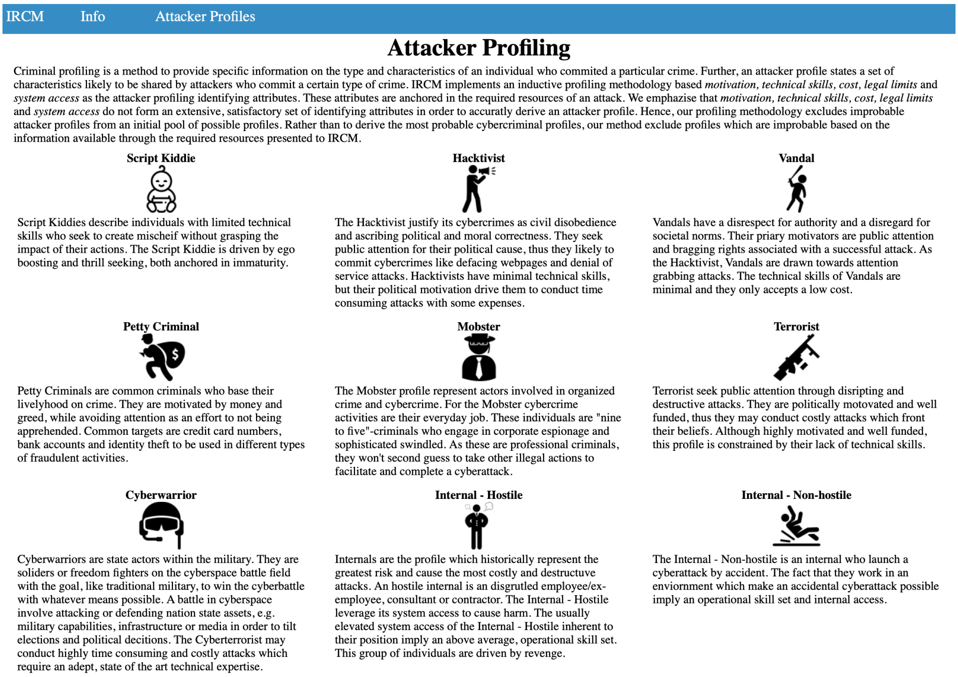
Minimum cost: 10

Confidence: 1.0

Figure D.8: Page displaying an already added resource alternative

D.2 IRCM Version 1

Below are screenshots of the IRCM v1 used in User Test 2, see 7.2.



Attacker Profiling

Criminal profiling is a method to provide specific information on the type and characteristics of an individual who commited a particular crime. Further, an attacker profile states a set of characteristics likely to be shared by attackers who commit a certain type of crime. IRCM implements an inductive profiling methodology based *motivation, technical skills, cost, legal limits* and *system access* as the attacker profiling identifying attributes. These attributes are anchored in the required resources of an attack. We emphazise that *motivation, technical skills, cost, legal limits* and *system access* do not form an extensive, satisfactory set of identifying attributes in order to accuratly derive an attacker profile. Hence, our profiling methodology excludes improbable attacker profiles from an initial pool of possible profiles. Rather than to derive the most probable cybercriminal profiles, our method exclude profiles which are improbable based on the information available through the required resources presented to IRCM.

Script Kiddie

Script Kiddies describe individuals with limited technical skills who seek to create mischeif without grasping the impact of their actions. The Script Kiddie is driven by ego boosting and thrill seeking, both anchored in immaturity.

Hacktivist

The Hacktivist justify its cybercrimes as civil disobedience and ascribing political and moral correctness. They seek public attention for their political cause, thus they likely to commit cybercrimes like defacing webpages and denial of service attacks. Hacktivists have minimal technical skills, but their political motivation drive them to conduct time consuming attacks with some expenses.

Vandal

Vandals have a disrespect for authority and a disregard for societal norms. Their priary motivators are public attention and bragging rights associated with a successful attack. As the Hacktivist, Vandals are drawn towards attention grabbing attacks. The technical skills of Vandals are minimal and they only accepts a low cost.

Petty Criminal

Petty Criminals are common criminals who base their livelyhood on crime. They are motivated by money and greed, while avoiding attention as an effort to not being apprehended. Common targets are credit card numbers, bank accounts and identity theft to be used in different types of fraudulent activities.

Mobster

The Mobster profile represent actors involved in organized crime and cybercrime. For the Mobster cybercrime activities are their everyday job. These individuals are "nine to five"-criminals who engage in corporate espionage and sophisticated swindled. As these are professional criminals, they won't second guess to take other illegal actions to facilitate and complete a cyberattack.

Terrorist

Terrorist seek public attention through disripting and destructive attacks. They are politically motovated and well funded, thus they may conduct costly attacks which front their beliefs. Although highly motivated and well funded, this profile is constrained by their lack of technical skills.

Cyberwarrior

Cyberwarriors are state actors within the military. They are soliders or freedom fighters on the cyberspace battle field with the goal, like traditional military, to win the cyberbattle with whatever means possible. A battle in cyberspace involve attacking or defending nation state assets, e.g. military capabilities, infrastructure or media in order to tilt elections and political decitions. The Cyberterrorist may conduct highly time consuming and costly attacks which require an adept, state of the art technical expertise.

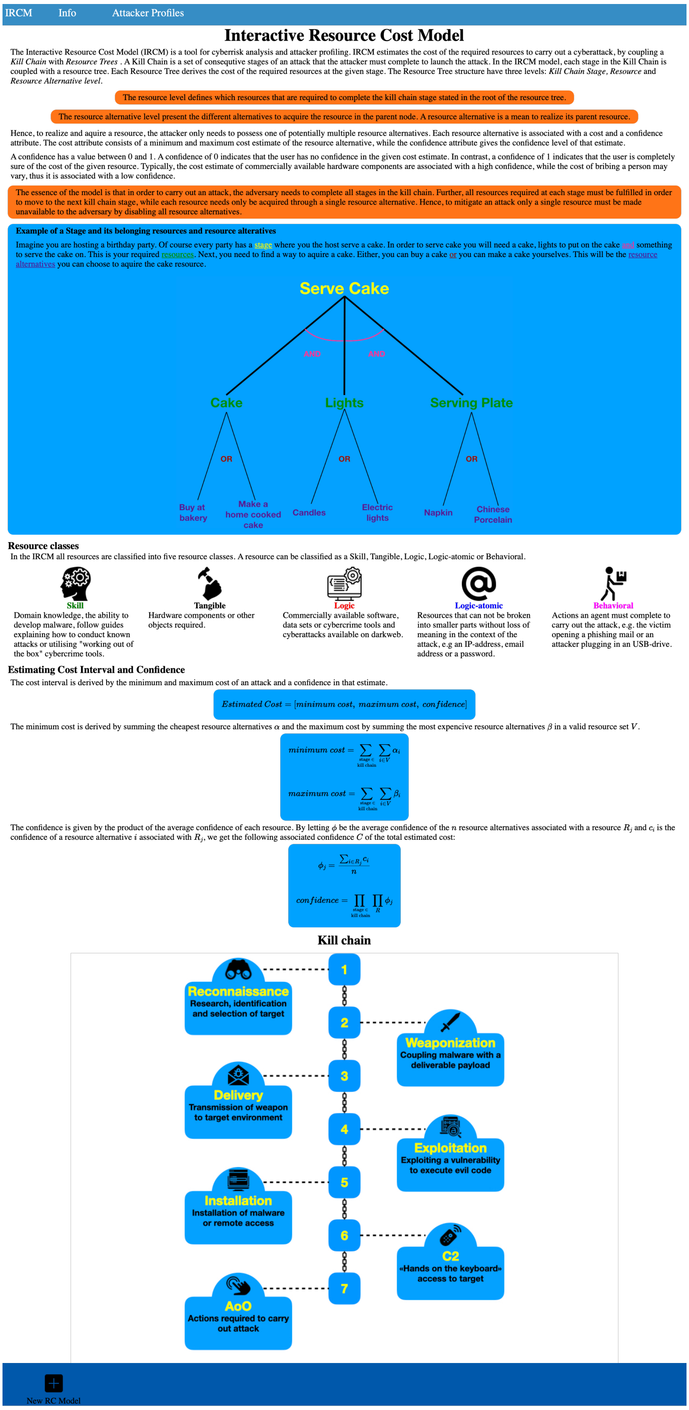
Internal - Hostile

Internals are the profile which historically represent the greatest risk and cause the most costly and destructuve attacks. An hostile internal is an disgrutled employee/ex-employee, consultant or contractor. The Internal - Hostile leverage its system access to cause harm. The usually elevated system access of the Internal - Hostile inherent to their position imply an above average, operational skill set. This group of individuals are driven by revenge.

Internal - Non-hostile

The Internal - Non-hostile is an internal who launch a cyberattack by accident. The fact that they work in an enviornment which make an accidental cyberattack possible imply an operational skill set and internal access.

Figure D.9: Information page on the different attacker profiles



Interactive Resource Cost Model

The Interactive Resource Cost Model (IRCM) is a tool for cyberrisk analysis and attacker profiling. IRCM estimates the cost of the required resources to carry out a cyberattack, by coupling a *Kill Chain* with *Resource Trees* . A Kill Chain is a set of consequtive stages of an attack that the attacker must complete to launch the attack. In the IRCM model, each stage in the Kill Chain is coupled with a resource tree. Each Resource Tree derives the cost of the required resources at the given stage. The Resource Tree structure have three levels: *Kill Chain Stage, Resource* and *Resource Alternative level*.

The resource level defines which resources that are required to complete the kill chain stage stated in the root of the resource tree.

The resource alternative level present the different alternatives to acquire the resource in the parent node. A resource alternative is a mean to realize its parent resource.

Hence, to realize and aquire a resource, the attacker only needs to possess one of potentially multiple resource alternatives. Each resource alternative is associated with a cost and a confidence attribute. The cost attribute consists of a minimum and maximum cost estimate of the resource alternative, while the confidence attribute gives the confidence level of that estimate.

A confidence has a value between 0 and 1. A confidence of 0 indicates that the user has no confidence in the given cost estimate. In contrast, a confidence of 1 indicates that the user is completely sure of the cost of the given resource. Typically, the cost estimate of commercially available hardware components are associated with a high confidence, while the cost of bribing a person may vary, thus it is associated with a low confidence.

The essence of the model is that in order to carry out an attack, the adversary needs to complete all stages in the kill chain. Further, all resources required at each stage must be fulfilled in order to move to the next kill chain stage, while each resource needs only be acquired through a single resource alternative. Hence, to mitigate an attack only a single resource must be made unavailable to the adversary by disabling all resource alternatives.

Example of a Stage and its belonging resources and resource alteratives

Imagine you are hosting a birthday party. Of course every party has a stage where you the host serve a cake. In order to serve cake you will need a cake, lights to put on the cake and something to serve the cake on. This is your required resources. Next, you need to find a way to aquire a cake. Either, you can buy a cake or you can make a cake yourselves. This will be the resource alternatives you can choose to aquire the cake resource.

Resource classes

In the IRCM all resources are classified into five resource classes. A resource can be classified as a Skill, Tangible, Logic, Logic-atomic or Behavioral.

Skill
Domain knowledge, the ability to develop malware, follow guides explaining how to conduct known attacks or utilising "working out of the box" cybercrime tools.

Tangible
Hardware components or other objects required.

Logic
Commercially available software, data sets or cybercrime tools and cyberattacks available on darkweb.

Logic-atomic
Resources that can not be broken into smaller parts without loss of meaning in the context of the attack, e.g an IP-address, email address or a password.

Behavioral
Actions an agent must complete to carry out the attack, e.g. the victim opening a phishing mail or an attacker plugging in an USB-drive.

Estimating Cost Interval and Confidence

The cost interval is derived by the minimum and maximum cost of an attack and a confidence in that estimate.

$$Estimated\ Cost = [minimum\ cost,\ maximum\ cost,\ confidence]$$

The minimum cost is derived by summing the cheapest resource alternatives α and the maximum cost by summing the most expencive resource alternatives β in a valid resource set V.

$$minimum\ cost = \sum_{\substack{\text{stage} \in \\ \text{kill chain}}} \sum_{i \in V} \alpha_i$$

$$maximum\ cost = \sum_{\substack{\text{stage} \in \\ \text{kill chain}}} \sum_{i \in V} \beta_i$$

The confidence is given by the product of the average confidence of each resource. By letting ϕ be the average confidence of the n resource alternatives associated with a resource R_j and c_i is the confidence of a resource alternative i associated with R_j, we get the following associated confidence C of the total estimated cost:

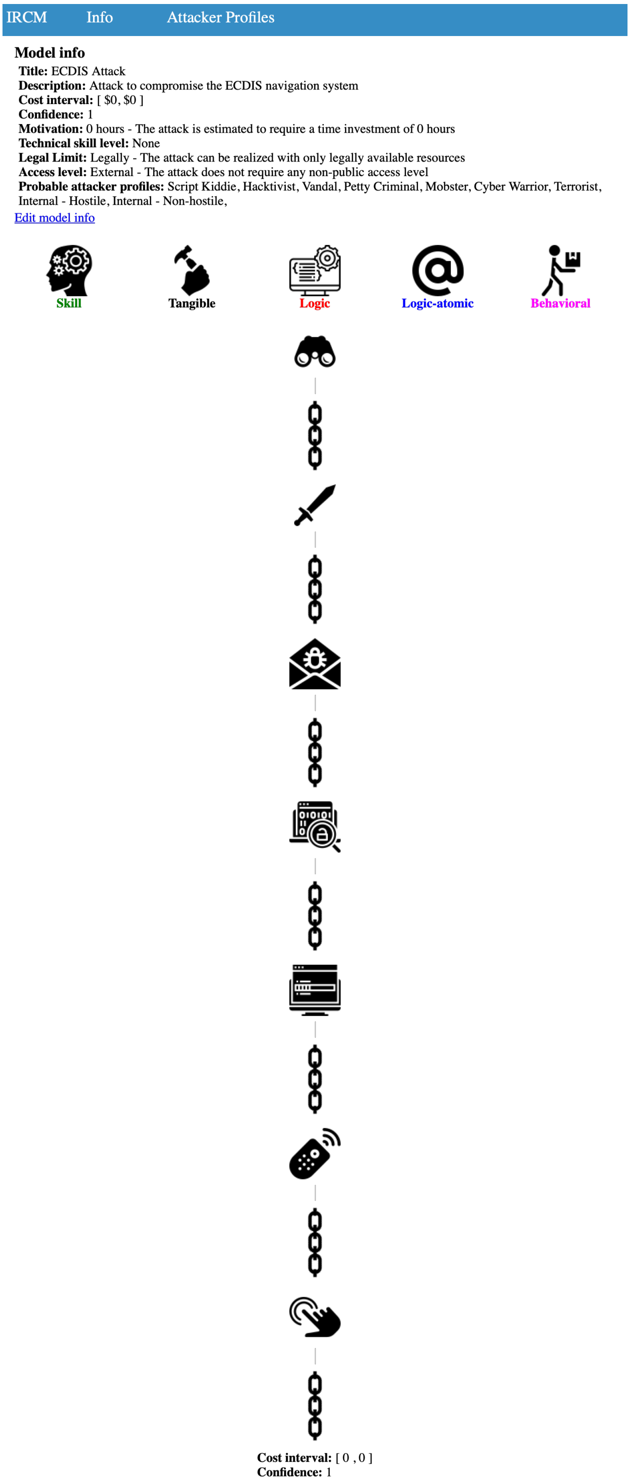
$$\phi_j = \frac{\sum_{i \in R_j} c_i}{n}$$

$$confidence = \prod_{\substack{\text{stage} \in \\ \text{kill chain}}} \prod_{R} \phi_j$$

Kill chain

1
Reconnaissance
Research, identification and selection of target
2
Weaponization
Coupling malware with a deliverable payload
3
Delivery
Transmission of weapon to target environment
4
Exploitation
Exploiting a vulnerability to execute evil code
5
Installation
Installation of malware or remote access
6
C2
«Hands on the keyboard» access to target
7
AoO
Actions required to carry out attack
New RC Model

Figure D.10: Info-page



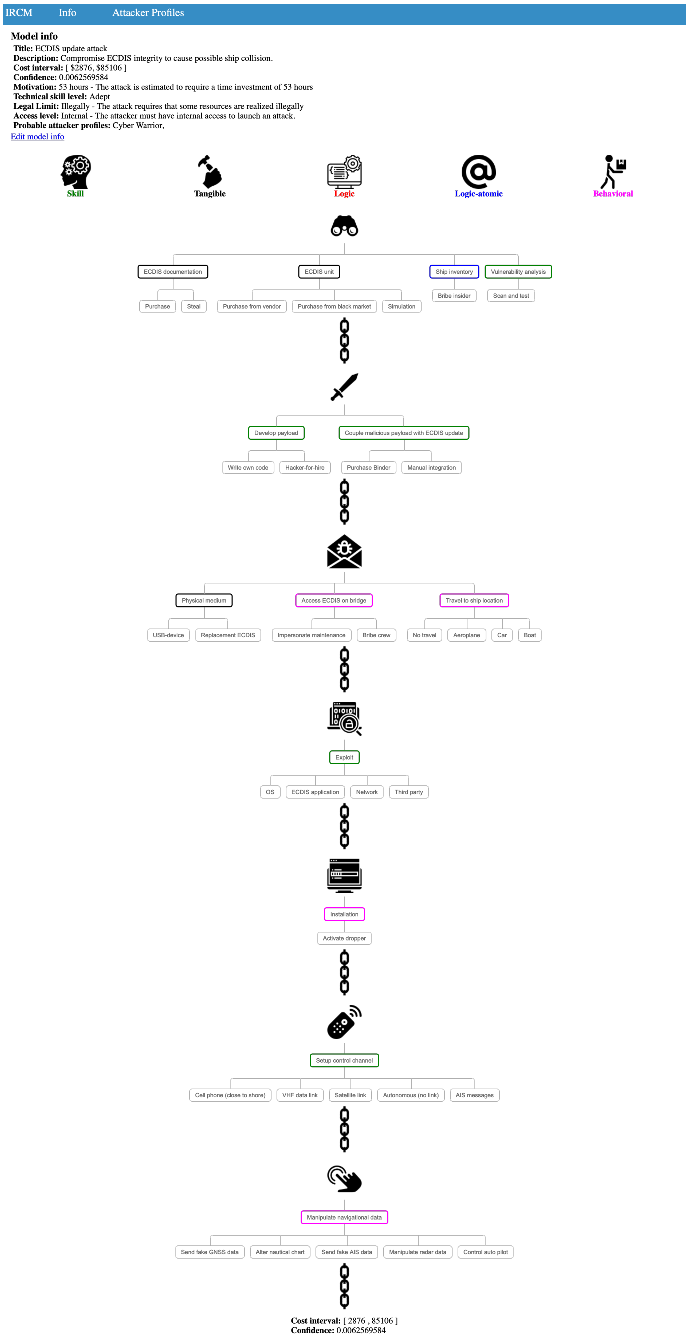
Model info

Title: ECDIS Attack
Description: Attack to compromise the ECDIS navigation system
Cost interval: [$0, $0]
Confidence: 1
Motivation: 0 hours - The attack is estimated to require a time investment of 0 hours
Technical skill level: None
Legal Limit: Legally - The attack can be realized with only legally available resources
Access level: External - The attack does not require any non-public access level
Probable attacker profiles: Script Kiddie, Hacktivist, Vandal, Petty Criminal, Mobster, Cyber Warrior, Terrorist, Internal - Hostile, Internal - Non-hostile,

Edit model info

Cost interval: [0 , 0]
Confidence: 1

Figure D.11: Page showing the RC Model before any resources are aded

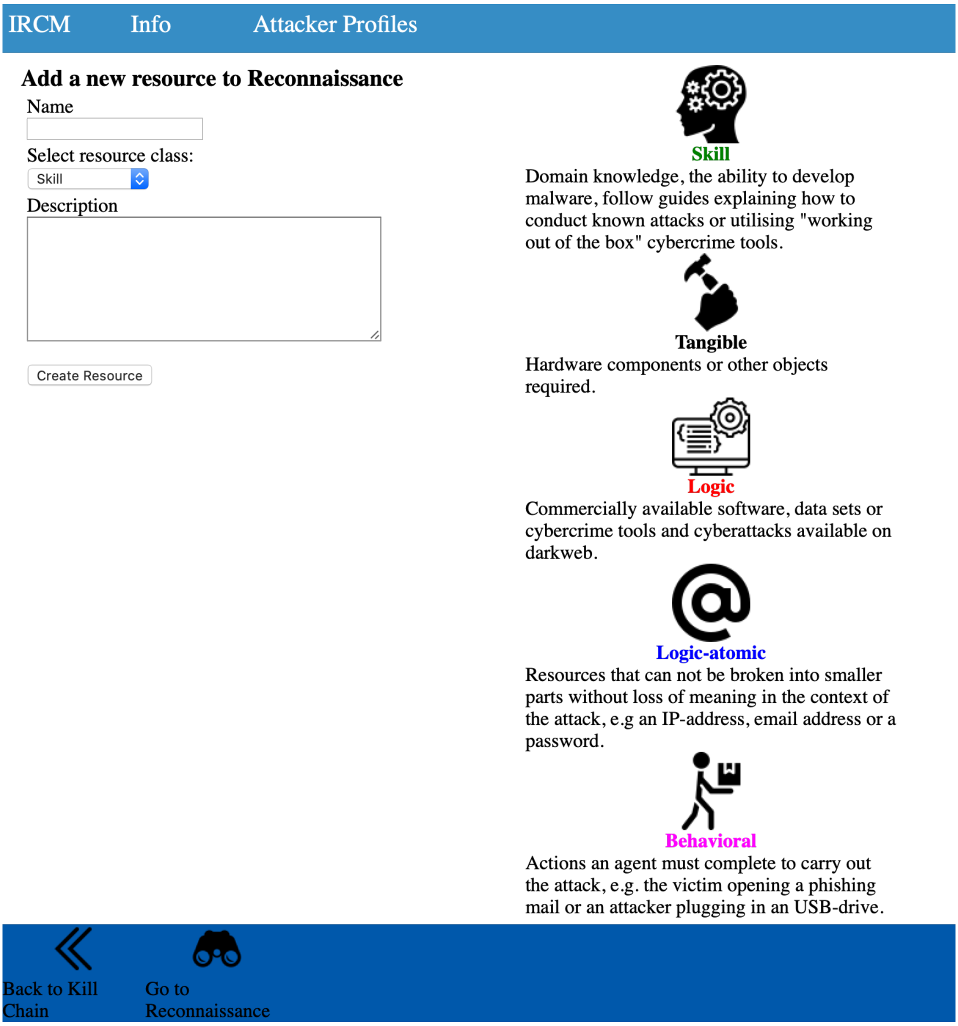


Model info

Title: ECDIS update attack
Description: Compromise ECDIS integrity to cause possible ship collision.
Cost interval: [$2876, $85106]
Confidence: 0.0062569584
Motivation: 53 hours - The attack is estimated to require a time investment of 53 hours
Technical skill level: Adept
Legal Limit: Illegally - The attack requires that some resources are realized illegally
Access level: Internal - The attacker must have internal access to launch an attack.
Probable attacker profiles: Cyber Warrior,

Cost interval: [2876 , 85106]
Confidence: 0.0062569584

Figure D.12: Populated RC Model modeling a ECDIS attack



Add a new resource to Reconnaissance

Name

Select resource class: Skill

Description

Skill
Domain knowledge, the ability to develop malware, follow guides explaining how to conduct known attacks or utilising "working out of the box" cybercrime tools.

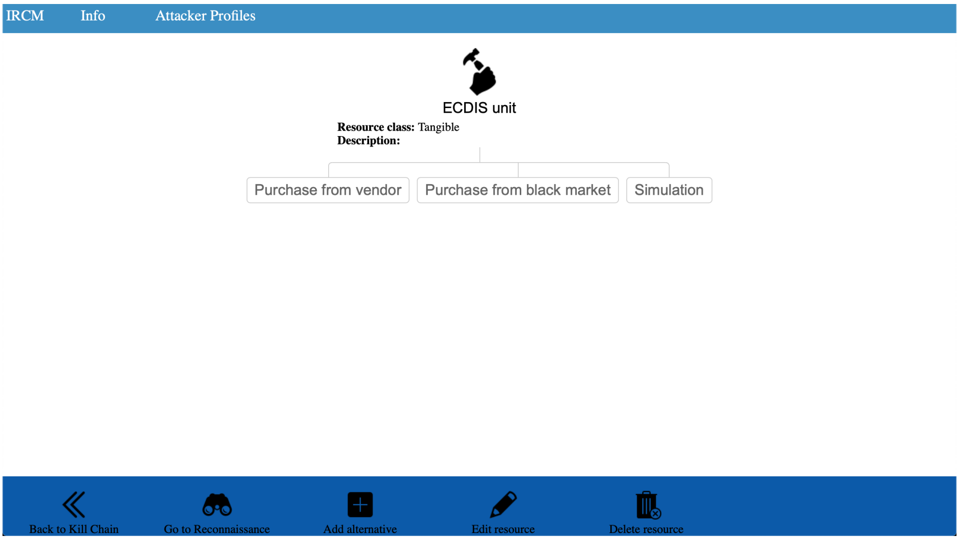
Tangible
Hardware components or other objects required.

Logic
Commercially available software, data sets or cybercrime tools and cyberattacks available on darkweb.

Logic-atomic
Resources that can not be broken into smaller parts without loss of meaning in the context of the attack, e.g an IP-address, email address or a password.

Behavioral
Actions an agent must complete to carry out the attack, e.g. the victim opening a phishing mail or an attacker plugging in an USB-drive.

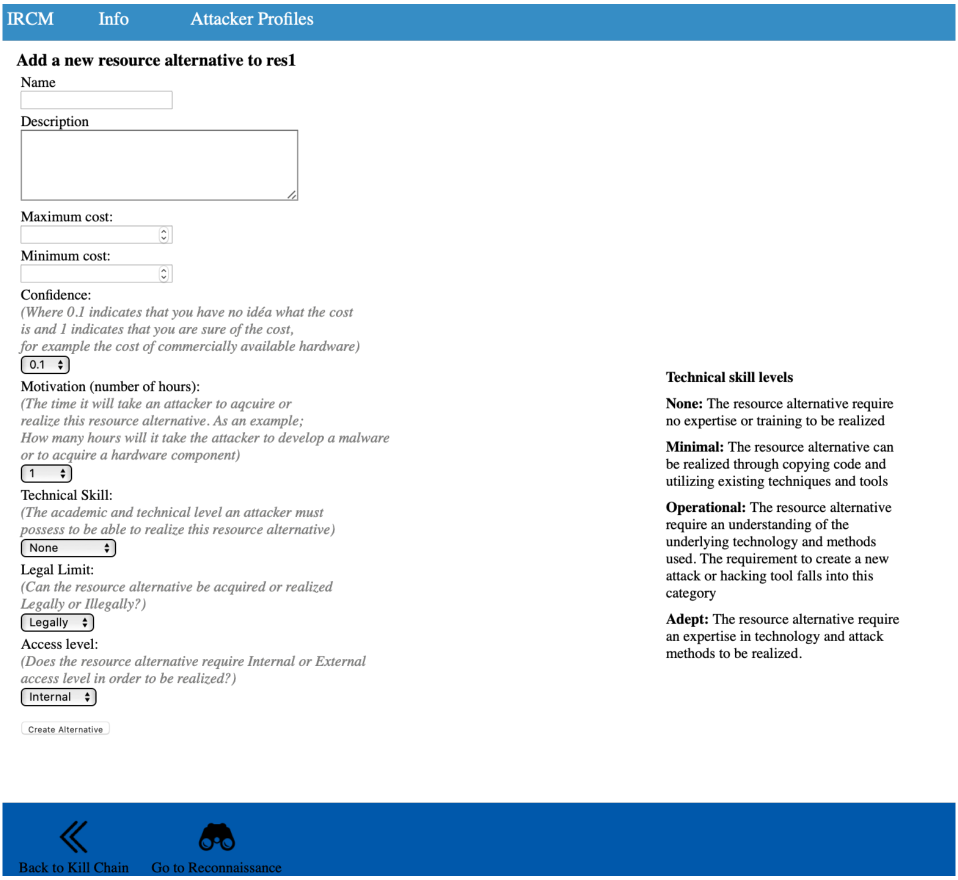
Figure D.13: Page for adding a resource. In order to add a resource the user first had to select the stage to add the resource by viewing page shown in D.4 before clicking ”Add resource”. The same page is used for editing a resource



ECDIS unit

Resource class: Tangible
Description:

Figure D.14: Page displaying an already added resource



Add a new resource alternative to res1

Name

Description

Maximum cost:

Minimum cost:

Confidence:
(Where 0.1 indicates that you have no idéa what the cost is and 1 indicates that you are sure of the cost, for example the cost of commercially available hardware)
0.1

Motivation (number of hours):
(The time it will take an attacker to aqcuire or realize this resource alternative. As an example; How many hours will it take the attacker to develop a malware or to acquire a hardware component)
1

Technical Skill:
(The academic and technical level an attacker must possess to be able to realize this resource alternative)
None

Legal Limit:
(Can the resource alternative be acquired or realized Legally or Illegally?)
Legally

Access level:
(Does the resource alternative require Internal or External access level in order to be realized?)
Internal

Technical skill levels

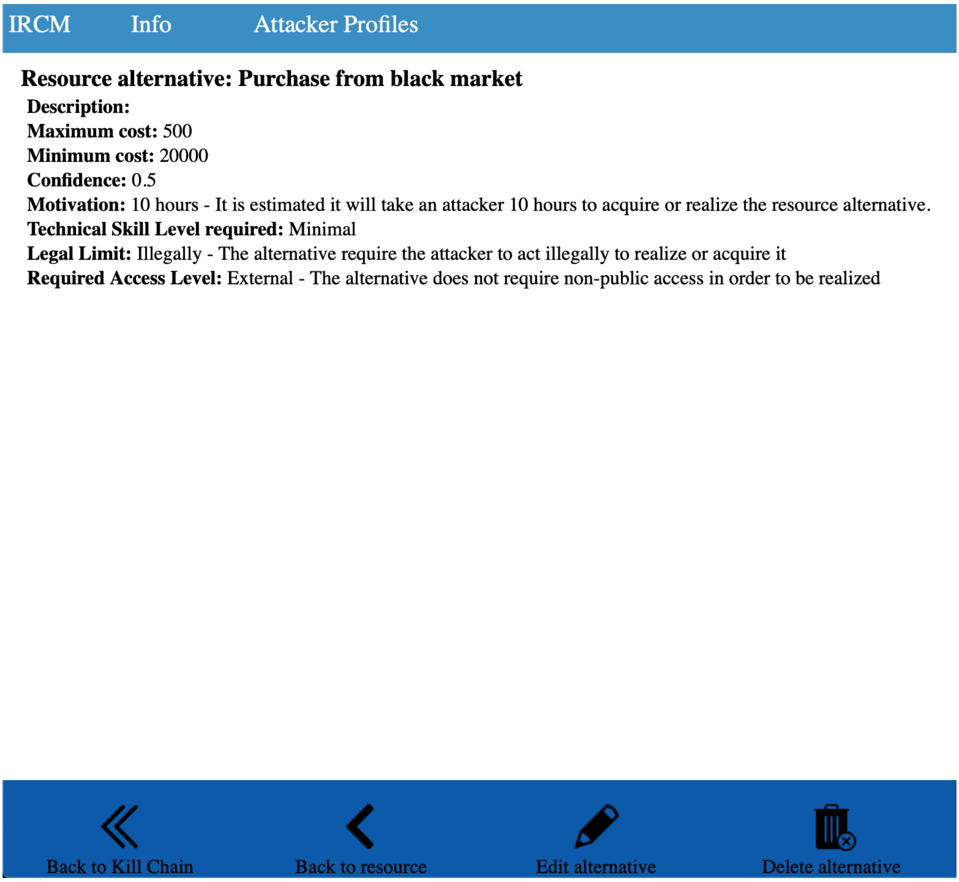
None: The resource alternative require no expertise or training to be realized

Minimal: The resource alternative can be realized through copying code and utilizing existing techniques and tools

Operational: The resource alternative require an understanding of the underlying technology and methods used. The requirement to create a new attack or hacking tool falls into this category

Adept: The resource alternative require an expertise in technology and attack methods to be realized.

Figure D.15: Page enabling the user to add a resource alternative. The user first selected the resource before being able to add a resource alternative by clicking ”Add alternative” shown is D.14. The same page is used for editing a resource alternative.



Resource alternative: Purchase from black market

Description:
Maximum cost: 500
Minimum cost: 20000
Confidence: 0.5
Motivation: 10 hours - It is estimated it will take an attacker 10 hours to acquire or realize the resource alternative.
Technical Skill Level required: Minimal
Legal Limit: Illegally - The alternative require the attacker to act illegally to realize or acquire it
Required Access Level: External - The alternative does not require non-public access in order to be realized

Figure D.16: Page displaying an already added resource alternative

Appendix E

IRCM v1 Development and Deployment Environment

IRCM v1 is developed from the Ruby on Rails (RoR) ”My First App”-scaffold and deployed through Github on the Heroku platform. RoR is a web-application framework in the *convention over configuration* paradigm, meaning that in RoR there is one right way to solve conceptual tasks, such as naming and file structure, decreasing the number of decisions a developer must make without loss of flexibility. Also, it empathizes the *Don't Repeat Yourself* (DRY) principle stating that ”Every piece of knowledge must have a single, unambiguous, authoritative representation within a system.” Further, RoR is a model–view–controller (MVC) framework, providing default structures for a database, a web service, and web pages. Starting development from the ”My First App”-scaffold, the software is developed following the default structures, the DRY principle and the RoR convention over configuration paradigm.

MVC frameworks divides program logic into three interconnected elements: models, views and controllers. This separate internal representations of information and how information in presented to the users.

The three MVC components have distinctive tasks:

- **Models** are the application's dynamic data structures, independent of the user interface. A model directly manages the data, logic and application rules.
- **Views** are any representation of information displayed to the user.
- **Controllers** accept and convert input to commands for the model or view.

Further, the MVC pattern defines the interaction between the three components. The controller responds to the user input requesting to manipulate model, which then updates the view that is perceived by the user who experience the response of its initial input. This flow of interaction is shown in Figure E.1.

The IRCM has a nested model where each entry in the resource alternative table is associated with a single entry in the resource table, further each resource is associated

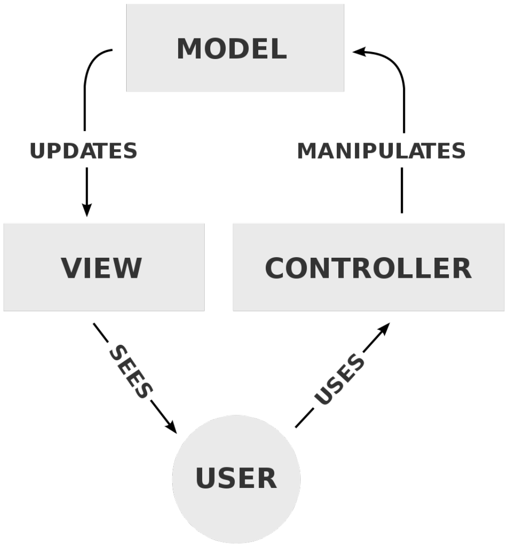


Figure E.1: Flow of interaction between a user and the MVC components

with a stage in the stage table and finally each stage is associated with a single entry in the RCM table. All tables, e.i RCM- , stage- , resource- and resource alternative table, have columns for table specific attributes such as name, resource class and resource alternative cost. The complete UML class diagram of the model is shown in Figure E.2.

According to RoR's convention over configuration each table has its own controller. The controller query and manipulate data in its corresponding model and control which model information that is available to the views routed to the controller. The nested model structure used in IRCM allow a controller to access data in models nested with its corresponding model. Here, the nested models allow the RCM controller to access data in the resource alternative model table.

The most complex controller in the IRCM application is the RCM controller which provide the information viewed in the RCM ”Show”-page, see Figure D.12. In addition to the conventional CRUD (create, read, update and delete) operations in each controller, the RCM controller aggregate the data from the resources and resource alternatives in order to derive the cost interval, confidence, required motivation, technical skills, legal limit and access level. The aggregated data is then fed to the cybercriminal exclusion rules implemented in the controller which provides the RCM view with the set of probable attacker profiles. The source code for the RCM controller can be inspected in Listing E at the end of this appendix.

The views in a RoR application have access to data in ”@variables” in the corresponding controller. These data are accessed and presented using the .html.erb extension. The extension allow programmers to mix HTML and Ruby code in the views files. The first extension states the format of the template, while the latter is the handler that is used to render the template. Further, the views are styled with Syntactically Awesome Style Sheets

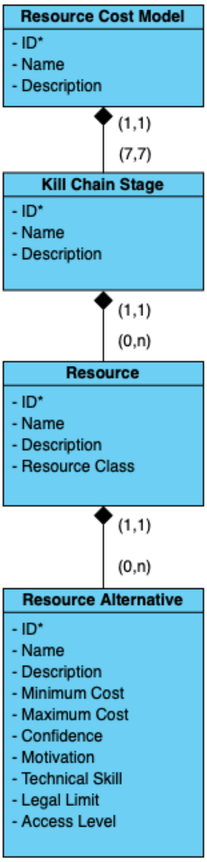


Figure E.2: UML class diagram of IRCM model

(.scss) style sheets in the application assets.

IRCM is deployed to Heroku using Git, a distributed source control management tool. Heroku is a platform as a service based on a managed container system with integrated data services. Heroku packages an app's code and dependencies into *dynos* - smart containers on a reliable, fully managed runtime environment that provide compute, memory, an OS, and an ephemeral filesystem - and the runtime keeps apps running without manual intervention. The main benefit of containers is the scalability, allowing apps to scale to any number of dyno containers to handle its resource demands.

For this study Heroku provided an easy and rapid setup of a remote, allowing for user testing via video link and shared screen features. IRCM is currently hosted on a sleeping dyno which require some wake up time on the first access, but afterwards it has similar performance to any dyno. IRCM can be accessed at:
`https://nameless-brushlands-94643.herokuapp.com`

```
class RcmodelsController < ApplicationController
  before_action :set_rcmodel, only:
            [:show, :edit, :update, :destroy]

  def index
    @rcmodels = Rcmodel.all
  end

  def show
    get_maxCost
    get_minCost
    get_confidence
    get_probable_attacker_profiles
  end

  def new
    @rcmodel = Rcmodel.new
  end

  def edit
  end


  def create
    @rcmodel = Rcmodel.new(rcmodel_params)

    if @rcmodel.save
      @rcmodel.stages.create(stage_params)

      redirect_to @rcmodel
    else
      render 'new'
    end
  end

  def update
    if @rcmodel.update(rcmodel_params)
      redirect_to @rcmodel
    else
      render 'edit'
    end
  end

  def destroy
```

```
    @rcmodel.destroy

    redirect_to rcmodels_path
  end

  private

    def set_rcmodel
      @rcmodel = Rcmodel.find(params[:id])
    end
    def rcmodel_params
      params.require(:rcmodel).permit(:title, :description)
    end

    def stage_params
      [
        {name: "Reconnaissance",
         description: "Research,␣identification␣and
␣␣␣␣␣␣␣␣␣selection␣of␣target"},
        {name: "Weaponization",
         description: "Coupling␣a␣malware␣with␣a
␣␣␣␣␣␣␣␣␣deliverable␣payload,␣e.g.␣an␣image,␣PDF
␣␣␣␣␣␣␣␣␣or␣Microsoft␣Office␣document"},
        {name: "Delivery",
         description: "Transmission␣of␣the␣weapon␣to␣the␣target
␣␣␣␣␣␣␣␣␣environment,␣e.g.␣an␣email␣attachment␣or␣USB-drive"},
        {name: "Exploitation",
         description: "Triggers␣malicious␣code.␣Ranges␣from
␣␣␣␣␣␣␣␣␣vulnerabilities␣or␣auto-executing␣features␣in␣host's
␣␣␣␣␣␣␣␣␣operating␣system␣to␣users␣triggering␣execution.
␣␣␣␣␣␣␣␣␣Also␣include␣factors␣triggering␣malicious␣code"},
        {name: "Installation",
         description: "Installation␣of␣malware␣or␣remote␣access"},
        {name: "C2",
         description: "Command␣and␣Controll␣(C2)
␣␣␣␣␣␣␣␣␣provides␣'hands␣on␣the␣keyboard'␣access
␣␣␣␣␣␣␣␣␣inside␣the␣target␣environment"},
        {name: "AO", description: "Actions␣on␣Objective␣(AO):
␣␣␣␣␣␣␣␣Actions␣required␣to␣acheive␣the␣goal␣of␣the␣attack"}
      ]
    end

    def get_maxCost
      @maxCost = 0
    @rcmodel.resources.each { |resource|
      if not resource.alternatives.pluck(:maxCost).max().nil?
        @maxCost += resource.alternatives.pluck(:maxCost).max()
      end
    }
  end

  def get_minCost
    @minCost = 0

    @rcmodel.resources.each { |resource|
      if not resource.alternatives.pluck(:minCost).min().nil?
        @minCost += resource.alternatives.pluck(:minCost).min()
      end
    }
  end

  def get_confidence
    @confidence = 1
    #confidenceArray = @rcmodel.alternatives.pluck(:confidence)
    #confidenceArray.each {|conf| @confidence *= conf }

    @rcmodel.resources.each { |resource|
      if not
      ( resource.alternatives.pluck(:confidence).sum().nil?
      ||
      resource.alternatives.pluck(:confidence).length() == 0 )
        avg_conf =
          resource.alternatives.pluck(:confidence).sum() /
          resource.alternatives.pluck(:confidence).length()
        @confidence *= avg_conf
      end

    }

    @confidence = @confidence.round(10)
  end

  def get_probable_attacker_profiles
    @initial_attacker_profiles = ['Script␣Kiddie',
    'Hacktivist', 'Vandal', 'Petty␣Criminal',
    'Mobster', 'Cyber␣Warrior', 'Terrorist',
    'Internal␣-␣Hostile', 'Internal␣-␣Non-hostile']

    get_profiles_excluded_by_motivation
    get_profiles_excluded_by_technical_skill
    get_profiles_excluded_by_cost
    get_profiles_excluded_by_limit
    get_profiles_excluded_by_access

    @probable_attacker_profiles =
      @initial_attacker_profiles
          - @excluded_by_motivation
          - @excluded_by_technical_skill
          - @excluded_by_cost - @excluded_by_limit
          - @excluded_by_access
  end

  def get_profiles_excluded_by_motivation
    get_minimum_motivation
    @excluded_by_motivation = []

    if @minimum_motivation > 0
      @excluded_by_motivation.append('Internal␣-␣Non-hostile')
    end
    if @minimum_motivation > 8
      @excluded_by_motivation.append('Script␣Kiddie')
    end
    if @minimum_motivation > 20
      @excluded_by_motivation.append('Vandal')
    end
    if @minimum_motivation > 40
      @excluded_by_motivation.append('Internal␣-␣Hostile')
      @excluded_by_motivation.append('Petty␣Criminal')
    end
    if @minimum_motivation > 100
      @excluded_by_motivation.append('Hacktivist')
      @excluded_by_motivation.append('Mobster')
    end
    if @minimum_motivation > 1000
      @excluded_by_motivation.append('Terrorist')
    end

  end

  def get_profiles_excluded_by_technical_skill
    get_required_technical_skill
    @excluded_by_technical_skill = []

    if @required_technical_skill == 'Adept'
```

```
        @excluded_by_technical_skill =
            ['Script␣Kiddie', 'Hacktivist',
                'Vandal', 'Petty␣Criminal',
                'Mobster', 'Terrorist',
                'Internal␣-␣Hostile',
                'Internal␣-␣Non-hostile']
      elsif @required_technical_skill == 'Operational'
        @excluded_by_technical_skill =
            ['Script␣Kiddie', 'Hacktivist',
                'Vandal', 'Terrorist']
      end
    end

    def get_profiles_excluded_by_cost
      get_minCost
      @excluded_by_cost = []

      if @minCost > 0
        @excluded_by_cost.append('Internal␣-␣Non-hostile')
      end
      if @minCost > 50
        @excluded_by_cost.append('Script␣Kiddie')
      end
      if @minCost > 100
        @excluded_by_cost.append('Vandal')
      end
      if @minCost > 500
        @excluded_by_cost.append('Internal␣-␣Hostile')
        @excluded_by_cost.append('Hacktivist')
      end
      if @minCost > 1000
        @excluded_by_cost.append('Petty␣Criminal')
      end
      if @minCost > 5000
        @excluded_by_cost.append('Mobster')
      end
      if @minCost > 10000
        @excluded_by_cost.append('Terrorist')
      end

    end

    def get_profiles_excluded_by_limit
      get_required_limit
      @excluded_by_limit = []
```

```
  if @required_limit == 'Illegally'
    @excluded_by_limit =
        ['Script␣Kiddie', 'Vandal', 'Internal␣-␣Hostile',
        'Internal␣-␣Non-hostile']
  end
end

def get_profiles_excluded_by_access
  get_required_access
  @excluded_by_access = []

  if @required_access == 'Internal'
    @excluded_by_access =
        ['Script␣Kiddie', 'Hacktivist', 'Vandal',
            'Mobster',
            'Petty␣Criminal', 'Terrorist']
  end
end

def get_minimum_motivation
  @minimum_motivation = 0

  @rcmodel.resources.each { |resource|
    alternative_motivations =
        resource.alternatives.pluck(:motivation)
    alternative_motivations.delete(nil)

    if not (alternative_motivations.nil? ||
        alternative_motivations.length() == 0 )
      @minimum_motivation += alternative_motivations.min()
    end
  }
end

def get_required_technical_skill
  @required_technical_skill = 'None'

  @rcmodel.resources.each { |resource|
    alternative_technical_skill_Array =
        resource.alternatives.pluck(:technical_skill)
    alternative_technical_skill_Array.delete(nil)

    if not (alternative_technical_skill_Array.nil? ||
        alternative_technical_skill_Array.length() == 0 )
      # When resources so far only require
      # ’None’ technical skills
      if @required_technical_skill == ’None’ &&
          !(alternative_technical_skill_Array.include? ’None’)

        if alternative_technical_skill_Array.include? ’Minimal’
          @required_technical_skill = ’Minimal’
        elsif alternative_technical_skill_Array.include?
        ’Operational’
          @required_technical_skill = ’Operational’
        elsif alternative_technical_skill_Array.include?
        ’Adept’
          @required_technical_skill = ’Adept’
        end

      # When one or more previous resourses require a
      # technical skill level of ”Minimal”
      elsif ( @required_technical_skill == ’Minimal’) &&
        ! ( (alternative_technical_skill_Array.include? ’None’) ||
           (alternative_technical_skill_Array.include? ’Minimal’))

        if alternative_technical_skill_Array.include?
        ’Operational’
          @required_technical_skill = ’Operational’
        elsif alternative_technical_skill_Array.include?
        ’Adept’
          @required_technical_skill = ’Adept’
        end

      # When one or more previous resourses require a
      # technical skill level of ”Operational”
      elsif @required_technical_skill == ’Operational’ &&
        ! ( (alternative_technical_skill_Array.include? ’None’) ||
          (alternative_technical_skill_Array.include? ’Minimal’) ||
          (alternative_technical_skill_Array.include?
          ’Operational’))

        if alternative_technical_skill_Array.include? ’Adept’
          @required_technical_skill = ’Adept’
        end

      end
    end
  }
```

```
    end

    def get_required_limit
      @required_limit = 'Legally'

      @rcmodel.resources.each { |resource|
        alternatives_limit = resource.alternatives.pluck(:limit)
        alternatives_limit.delete(nil)

        if !(alternatives_limit.nil? ||
                alternatives_limit.length() == 0 )
          if ! alternatives_limit.include? 'Legally'
            @required_limit = 'Illegally'
            break
          end
        end
      }


    end

    def get_required_access
      @required_access = 'External'

      @rcmodel.resources.each{ |resource|
        alternatives_access = resource.alternatives.pluck(:access)
        alternatives_access.delete(nil)

        if ! ( alternatives_access.nil? ||
                alternatives_access.length() == 0 )
          if ! alternatives_access.include? 'External'
            @required_access = 'Internal'
            break
          end
        end
      }
    end

end
```

Kristian Haga

Breaking the Cyber Kill Chain

NTNU
Norwegian University of
Science and Technology

SINTEF