MAEVA: A Framework for Attack Incentive Analysis with Application to Game Theoretic Security Assessment

Louai Maghrabi
Department of Cybersecurity
School of Engineering, Computing & Informatics
Dar Al-Hekma University
Jeddah, Saudi Arabia
LMaghrabi@DAH.edu.sa

Eckhard Pfluegel
School of Computer Science & Mathematics
Faculty of Science, Engineering & Computing
Kingston University
London, United Kingdom
E.Pfluegel@Kingston.ac.uk

***Abstract*—This paper is concerned with the risk assessment of cyber security attacks on an organisation. We develop the novel attack incentive analysis framework Motive, Ability, Exploitability, Visibility and Attractiveness (MAEVA) based on taking into account a multiplicative function of the attacker's anticipated attack effort and expected reward. We argue that our approach can complement and enhance the standard approach based on estimating risk as a function of attack likelihood and impact on the organisation. We then present an application of our framework to game-theoretic risk assessment, illustrating how it can be used to inform the modelling of attacker-defender scenarios using complete information games. This helps to establish more realistic game-theoretical modelling of security assessment scenarios for practical use.**

***Index Terms*—Cybersecurity, risk assessment, game theory, security games, Nash equilibrium analysis**

I. Introduction

With the advancement and continuous growth of the digitally connected world through the Internet, cyber security has become a matter of global interest and importance to governments and private organisations to ensure achieving the major security requirements of Confidentiality, Integrity and Availability (CIA) of critical assets. To put this into some context, for example, a large organisation, such as SolarWinds recently had a data breach through hidden malicious code inserted into widely-used SolarWinds software, without being detected for several months. The attack gave adversaries access to systems of multiple U.S. government departments, including the Energy Departments nuclear arsenal. In another recent incident, Garmin, makers of Global Positioning System (GPS) devices, smart wearable devices and aviation technology, suffered a ransomware attack that brought down its own systems affecting the availability of data [1].

Security incidents of such high severity highlight the importance of security controls and mitigation techniques, and most governments and organisations nowadays have developed some form of strategies to categorise risks, apply vulnerability controls and mitigate threats in order to protect critical assets. National and international standards exist, to recommend formal frameworks and security management methodologies. *Security management* refers to a collection of activities that seek to, in the most general sense, the identification, assessment, analysis, establishment and evaluation of the security of a system or an organisation. This process can be carried out in different contexts such as information security, network security, system application or software security or nowadays generally in cyber security. Managing the security of an organisation can reduce the risk of running unexpected costs, help with standardising security practices, and show effective compliance with legislation and regulatory policies.

Risk management is the risk-based, top-down approach of security management. According to the National Institute of Standard and Technology (NIST), risk management is established as a risk context by producing a risk management strategy on how to identify, assess, respond, mitigate and monitor risks within an organisational context [2]. Generally, the following are typical risk management activities:

- Decide on how to implement a protection strategy and design risk mitigation plans by developing an action plan;
- Implement the detailed action plan;
- Monitor the action plans for schedule and effectiveness;
- Control variations in plan execution by taking appropriate corrective actions.

In this paper, we are studying the fundamental problem of how to compute the risk that an organisation faces from external attacks, and how to respond to it. According to [3], there are many approaches to assess risks. Risks can be assessed through qualitative or quantitative approaches, with underlying mathematical models of various degrees of complexity. Fundamentally, risk assessment attempts to measure the impact of an attack on an asset, mitigated by the probability (likelihood) that the attack will occur. In [4], the additional difficulty of a large (and ever-growing) attack surface of typical organisations and their assets, and the fact that risk can be seen as a map with different values at each point of the enterprise attack surface, is reported. Risk is seen as a

 ISBN: 978-1-61208-862-4

function of vulnerabilities in the system, their exposure to an attacker, the presence of active (relevant) threats, the existence of mitigating controls and the impact on the organisation. In this paper, we are interested in the risk assessment stage. We assume that prior to this step, critical assets and their security requirements were identified and that the above-stated relevant attack surface parameters are known.

This paper presents two contributions. The first contribution is a novel framework for risk assessment of cyber security attacks on an organisation. The framework is based on analysing the incentive an adversary may have to attack the organisation when weighing up the potential gain from the attack and the effort it takes to breach the system. We argue that this point of view, which is fundamentally different to that taken in traditional risk assessment, can complement and enhance the standard approach based on estimating risk as a function of attack likelihood and impact on the organisation. The second contribution is an application of this framework to game-theoretic risk assessment. We show that our framework is very convenient when wishing to inform the design of complete information games, modelling attacker-defender scenarios. It is hence a natural first step an organisation can take to prepare a game-theoretic risk assessment, and to reap the benefits from this approach which might have advantages compared to standard risk assessment.

This paper is organised as follows. Section II reviews modern risk assessment methodologies and formulates the main research question. In the subsequent section, the novel framework is introduced. Section IV proposes the application to game theory. The last section is the conclusion.

II. Security Assessment Based on Risk Analysis

The fundamental problem of how to compute the risk that an organisation faces from security attacks is the subject of numerous security risk assessment methodologies. In this section, we will recall the principles of risk computation and its challenges, review some popular mature security assessment frameworks, and discuss how they can help with the task of attack likelihood assessment and impact analysis.

A. The Challenges of Risk Computation

Using formal notation, the risk R can be expressed as an expected impact I, computed using the equation:

$$R = p \cdot I \tag{1}$$

where p is the probability of an attack occurring, often referred to as *attack likelihood*. From this equation, one can see that the problem now is to quantify and compute p and I and the difficulty is to perform a realistic estimate of these variables. Likelihood assessment is the process of establishing an estimate for p [5]. However, as pointed out in Tripwire [6], likelihood assessment appears, in general, to be a challenging and elusive task. Informally, the impact I is the overall damage that the targeted asset owner suffers from, this includes any indirect cost to the organisation such as a loss of reputation or business revenues. Impact is a central concept in the various security assessment frameworks, although it is defined slightly differently. This will be explored further in the following sections.

B. NIST

NIST Risk Management Framework (RMF) is a popular and detailed framework. Quoting [2], it states that *"the level of **impact** from a **threat** event is the magnitude of harm that can be expected to result from the consequences of unauthorised disclosure of information, unauthorised modification of information, unauthorised destruction of information, or loss of information or information system availability."* In other words, the impact from an attack on an asset is the degree of harm that affects the security requirements of confidentiality, integrity and availability for an asset. In this definition, the impact is created by a threat event, in line with the risk-based approach explained earlier. It is assumed that one is able to determine the attack likelihood. This leads to a table containing risk response actions, such as defending critical assets, recovering from an attack, planning for defense or choosing not to respond at all [2]. An appropriate response action is then determined by indexing the table rows with attack probabilities using qualitative metrics (low, medium or high) and its columns with a measure for the impact (minor, moderate or major) of the attack on the asset, or more generally, the organisation as a whole. This table, referred to as *Risk Response Matrix* (RRM) in this paper, is illustrated in Figure 1.

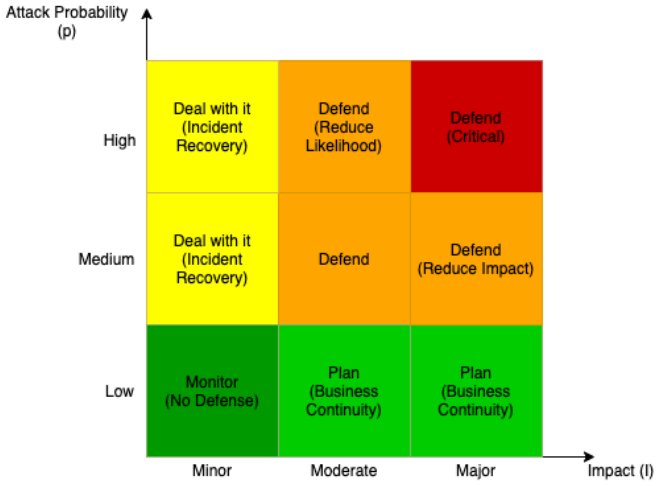

Fig. 1. Risk Response Matrix (RRM) [2]

C. OCTAVE

The Operationally Critical Threat and Vulnerability Evaluation (OCTAVE) framework [7] can be used to relate impact to both threats and vulnerabilities: *"All aspects of risk (assets, **threats**, **vulnerabilities**, and organisational **impact**) are factored into decision making, enabling an organisation to match a practice-based protection strategy to its security risks."* This

framework does not explicitly link the analysis of risk to the probability of an attack occurring. Instead, it informs the analysis based on threat profiling, enhanced by impact statements, leading to risk profiles. OCTAVE recommends at least looking at the following impact areas: safety, health, productivity, reputation, financial and fines. The analysis is done in a qualitative manner, but approximated scores could be derived from this.

D. CVSS

The Common Vulnerability Scoring System (CVSS) primarily focuses on software vulnerabilities, and the assessment of their severity. The idea is to provide a *base score* $\mu_B(v)$ for a CVE-indexed vulnerability v based on open criteria, and to make the score publicly available on the National Vulnerability Database (NVD) website [8]. This overall base score is further refined using an *impact score* $\mu_I(v)$ and an *exploitability score* $\mu_E(v)$. Quoting from [9], *"...the **impact** metrics reflect the direct consequence of a successful **exploit**, and represent the consequence to the thing that suffers the **impact**, which is referred to, formally, as the impacted component."* CVSS formulates the impact as the direct damage to an asset through an exploited vulnerability. In the context of risk assessment, we can hence use $\mu_I(v)$ for impact computation for the subset of suitable assets. It is less clear, how this could help with attack likelihood computation.

E. STRIDE and DREAD

STRIDE [10] was introduced in 1999 by Microsoft as a threat profiling scheme for categorizing potential threats according to their impact on common security requirements. The STRIDE acronym is formed from the first letter of each of the following categories, which cover a fairly complete range of threats when considering the original context of secure application development:

1) **S**poofing identity: illegally accessing and using another user's authentication credentials.
2) **T**ampering with data: malicious modification, fabrication or deletion of data.
3) **R**epudiation: the denial of having performed an action, in an environment lacking the capability to prove otherwise.
4) **I**nformation disclosure: exposure of information to individuals who are not authorised to have access to it.
5) **D**enial of service (DoS): an attack that interrupts the availability of a service to valid users.
6) **E**levation of privilege: an unauthorized or unprivileged user gains privileged access and thereby has sufficient access to compromise or destroy the entire asset or system.

Hence, risk assessment with STRIDE consists of eliciting threats using the approach above, followed by a rating system in order to rank threats by their criticality. This can be done using the less well-known DREAD [11] approach, based on the following key categories:

1) **D**amage potential: the degree of the potential damage a specific threat can inflict on an asset.
2) **R**eproducibility: this gives an understanding of the level of complexity of the threat, by assessing how easily it can be replicated by different adversaries.
3) **E**xploitability: this aims to quantify how easy is it for an attacker to succeed in exploiting the vulnerability targeted by the threat.
4) **A**ffected users: an estimate of the number of affected users in the aftermath of the attack.
5) **D**iscoverability: How difficult is it to discover vulnerabilities in the system, targeted by the threat.

By inspecting all of these DREAD categories and adding up individual scores, a risk rating is determined for each threat and the vulnerabilities affected by it.

III. A Framework for Attack Incentive Analysis

Assessing and responding to risk based on estimating attack likelihood and impact, and deciding on suitable response actions by forming and inspecting the corresponding risk response matrix seems natural and intuitive. While it is indeed a mainstream approach used in popular mature and standard security assessment frameworks as reviewed in the previous section, it is somewhat self-centric and might only lead to a limited view of the external threat and attack landscape. In particular, it fails to take into account the attacker's capabilities and perspective, in terms of his or her underlying motivation of the attack, knowledge of the target and its vulnerabilities, as well as the expected benefits gained. In this section, an alternative approach for informing risk assessment is devised, based on estimating the incentive to attack, that the adversary may have. While it seems very reasonable to assume that an informed attacker would wish to follow this framework, we will also suggest that the framework could be useful for the organisation that might be targeted by the attacker, as an alternative security assessment approach. This aspect will be further explored in the discussion part of this section.

A. Attack Incentive Matrix

An attacker is mainly motivated by the anticipated reward from the attack, which will be referred to as the *gain* in this paper, denoted by G. This gain however will be diminished by the *effort* he or she has to invest in order to implement the attack. This effort, denoted by e, is spent by exploiting (technical) vulnerabilities and breaching cyber security defences. Overall, the *attack incentive* A can be computed as

$$A = e^{-1} \cdot G. \tag{2}$$

Although this equation is simple and might not accurately reflect a potentially more complex inter-dependency of the involved parameters in real scenarios, we want to maintain a degree of simplicity which is comparable to that in the risk computation formula (1).

This idea leads to our proposed *Attack Incentive Matrix* (AIM) depicted in Figure 2, describing possible actions that the attacker might take, depending on the attacker's expected

gain (low, medium or high) and effort (minor, moderate or major) reference by the rows and columns of the matrix. We propose to specify the following actions: plan to attack (monitor target), information gathering (reconnaissance), and attack the target.

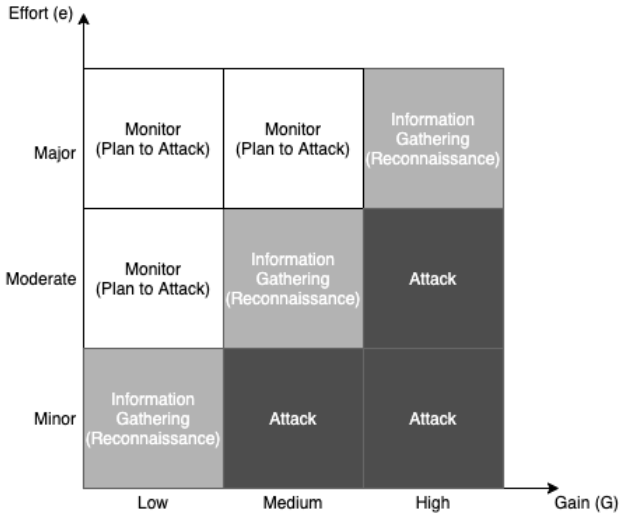

Fig. 2. Attack Incentive Matrix (AIM)

B. Proposed Framework: MAEVA

In this section, we present our MAEVA framework as guidance for computing the attack incentive A, based on estimating the effort e and the gain G. This proposed framework has several key characteristics partially inspired by considerations already used in the security assessment methodologies reviewed in the previous section, however in a different context. Our framework recommends considering the following categories when trying to estimate the required parameters, guided by following the MAEVA mnemonic:

1) **M**otive: the underlying reason for attacking the victim. This could be for the purposes of financial gains, revenge, personal satisfaction or thrill, or simply with the intention of creating damage. From a psychological point of view, the attacker's motive might affect the perceived gain, as well as the appreciation of the effort required.
2) **A**bility: the capability of the attacker to invest in resources for implementing the attack, as well as the technical knowledge available for breaching cyber security controls. A strong ability will make it easier to spend effort on the attack, and subjectively reducing the perceived value of e.
3) **E**xploitability: the ease by which the system can be penetrated, through exploiting a vulnerability. It would be reasonable to expect exploitability and effort to be inversely related in a proportional manner. This category could be explored similarly as in the CVSS exploitability score, taking into account possible attack vectors and attack complexities, as well as the required privileges and interaction with users, however, the discussion should not be restricted to software vulnerabilities alone.
4) **V**isibility of target: how prominent is the target, for example, does it have a popular website or brand name, does it have a large user base? Great visibility might promise a big gain, in the eyes of the attacker.
5) **A**ttractiveness of target: from the point of view of the attacker, how attractive is the target? This is linked to how much gain the attacker would estimate from achieving through the attack, and will strongly depend on the specific motive, as discussed in the first category.

While the MAEVA framework is intended to be used by the risk assessing organisation, it is an attacker-aware framework and the main assumption of its use is that an attacker would find it very natural to follow the same methodology in order to have a more systematic way to locate specific points within the AIM matrix in a given scenario, and use this as a guide for the decision to attack or not.

C. Discussion

We have reviewed the security assessment approach based on computing risk and introduced an alternative framework for modelling the attack incentive. Both approaches bring challenges in terms of achieving precise estimates for realistic results in practical scenarios. This will be briefly discussed and the advantages of combining both approaches outlined in this section.

In the risk matrix approach, the parameter that is more challenging to estimate is the attack likelihood p, as it depends a lot on external factors outside of our control. When trying to model attack incentive matrices, the difficult parameter is the effort e, since this has to be viewed as a relative quantity, depending on the capabilities of the attacker. Both approaches are complementary and if we use both, we can develop a better understanding of the risk that the organisation faces from an impeding cyber security attack.

By taking into account both perspectives (attacker, defender), a good understanding of the impact I can be developed by comparing it with the gain G. A discrepancy might reveal the need for correcting any of those two parameters. Furthermore, the attack likelihood p would be closely related to the attack incentive A, and this can help with computing R. As the effort e will depend, amongst other things, on the organisation's willingness to apply a security control, in other words, the perceived risk R, it might be necessary to adapt the estimate for A. After several iterations of estimations and adaptions, a final model should be obtained. We argue that the resulting figures are much more reliable and realistic, than those obtained without using MAEVA.

IV. Application to Game Theory

In this section, we will show that the RRM together with the AIM approach can be used naturally when modelling a non-cooperative two-player non-zero-sum complete information game, which is a specific type of security game useful for game-theoretic risk assessment. A complete information game

 ISBN: 978-1-61208-862-4

means that each player knows the strategies and payoffs of the other player in the game, but not necessarily the actions. For more background information on security games, we refer to [12].

A. Game Description

We are concerned with a single-target security game $G(\mathcal{D}, \mathcal{A})$ where the main focus is on a single asset that has an exploitable vulnerability. Our simple game comprises of two players: an attacker $\mathcal{A}$ and a defender $\mathcal{D}$ where each player has their own strategies as illustrated in Table I. The rows corresponds to the strategies available to the defender: $S_{\mathcal{D}} = \{\text{defend, not defend}\} = \{s_d, s_{-d}\}$, and the columns indicates the attacker's strategies: $S_{\mathcal{A}} = \{\text{attack, not attack}\} = \{s_a, s_{-a}\}$. Moreover, there is a payoff function (e.g., cost and benefit) that each player will incur depending on their chosen strategy: $c_{\mathcal{D}}$ is the defence cost, I is the defender's loss (impact) from an attack. By $c_{\mathcal{A}}$ we denote the attacker's cost, and G is the gain (benefit) of the attacker from an attack. Note that we have used notations that are compatible with the previous sections. The following natural assumptions [13] are usually made for this type of security game: the *Principle of Adequate Protection* prescribes that defence costs must not exceed potential losses: $c_{\mathcal{D}} < I$, and the *Principle of Easiest Attack* states that the attacker prefers to keep his or her cost for attacking bounded by the expected gain: $c_{\mathcal{A}} < G$. The game is described using its payoff matrix, which specifies its strategic normal form:

TABLE I
PAYOFF MATRIX FOR $G(\mathcal{D}, \mathcal{A})$

$\mathcal{D} \downarrow \mathcal{A} \rightarrow$	s_a	s_{-a}
s_d	$-c_{\mathcal{D}}$, $-c_{\mathcal{A}}$	$-c_{\mathcal{D}}$, 0
s_{-d}	$-I$, $G - c_{\mathcal{A}}$	0, 0

B. Game Analysis

When using a so-called *Nash Equilibrium strategy*, none of the players will have the incentive in deviating unilaterally from this strategy as this will reduce his or her expected utility. The following results are well-known properties of security games such as the game G, c.f. [12].

Theorem 1. *The security game $G(\mathcal{D}, \mathcal{A})$ has no pure Nash Equilibrium.*

Proof. By inspecting the payoff matrix of the game. □

Theorem 2. *A mixed Nash Equilibrium strategy pair $(x^*_{\mathcal{D}}, x^*_{\mathcal{A}})$ is obtained, where $q^* = 1 - c_{\mathcal{A}}/G$ and $p^* = c_{\mathcal{D}}/I$ are the probability of defense and attack respectively.*

Proof. Following Nash, as further detailed in [12]. □

In the context of security assessment, the outcomes of the game analysis have the following implications:

- Due to the lack of a pure equilibrium solution, there is no clear-cut decision whether to defend or not, as there is a dilemma between the conflicting non-cooperating players of the game.
- The mixed equilibrium solution can be interpreted as a means to compute risk, by interpreting the mixed strategy of the attacker as a probability value: $R = p^* \cdot I$.

Hence, a more systematic and theoretically justified way to compute risk can be achieved, based on game theory.

C. MAEVA Application

As we have seen, under the assumption of complete information about the strategies available to both players, the use of game theory improves the traditional risk assessment approaches as it combines both the non-cooperative nature of the defender and the attacker. Before the game can be solved, it needs to be specified in terms of the precise values for the payoff functions, and Table I reveals that the MAEVA framework can be used to determine (an estimate for) G. The parameter $c_{\mathcal{D}}$ is effectively the *defense budget* of the organisation and $c_{\mathcal{A}}$ can be related to the attacker's effort e. Hence, in a natural way, both the RRM and AIM methodologies provide the input parameters for the game. The analysis of the game based on computing the Nash equilibrium will then result in the desired risk value, following the computation as presented in the previous section.

V. CONCLUSION

In this paper, we have proposed a new framework entitled MAEVA, for analysing the attack incentive of a cyber security adversary of an organisation. Furthermore, we have shown how to use this framework in combination with traditional risk analysis, in order to achieve a more refined strategy to assess typical risk-related parameters such as attack likelihood and impact. We have also demonstrated that the framework is useful as preparation of game-theoretic modelling of risk assessment. To our knowledge, our framework constitutes a novel approach and we recommend using it as a practical methodology for any organisation wishing to assess risk, perhaps in combination with other mainstream methods.

The next step for this research would be an implementation of a real scenario, and a detailed evaluative comparison with existing approaches. For example, an organisation could review their information assets, apply both the RRM and AIM, and compare the resulting parameters. It would be interesting to relate this to historical information about cyber security incidents that happened in the past at this organisation, or in its sector. Ideally, we would expect an advantage resulting from the dual use of these frameworks, in terms of obtaining more realistic risk estimates. While not being the main focus of this paper, another interesting aspect that deserves further attention is to more deeply explore the link between traditional and game-theoretical security assessment. The authors believe that risk assessment modelling using game theory would have numerous advantages and that it should be considered for use in future versions of mainstream security assessment methodologies.

 ISBN: 978-1-61208-862-4

REFERENCES

[1] M. Gostovnikas, "The 9 worst recent data breaches of 2020," https://auth0.com/blog/the-nine-worst-recent-data-breaches-of-2020, January 2021. [Online].

[2] NIST, "NIST SP 800-30: Guide for conducting risk assessments," NIST, Tech. Rep., September 2012. [Online]. Available: http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf{\%}5Cnhttp://csrc.nist.gov/publications/PubsSPs.html{\%}5Cnhttp://dx.doi.org/10.6028/NIST.SP.800-30r1

[3] S. Derakhshandeh and N. Mikaeilvand, ""New framework for comparing information security risk assessment methodologies"," *Australian Journal of Basic and Applied Sciences*, vol. 5, no. 9, pp. pp. 160–166, 2011.

[4] G. Banga, ""Why is cybersecurity not a human-scale problem anymore?"," *Commun. ACM*, vol. 63, no. 4, p. pp. 3034, March 2020. [Online]. Available: https://doi.org/10.1145/3347144

[5] W. Kanoun, F. Cuppens-Boulahia, N. Cuppens, S. Dubus, and A. Martin, ""Success likelihood of ongoing attacks for intrusion detection and response systems"," in *2009 International Conference on Computational Science and Engineering*, vol. 3, 2009, pp. pp. 83–91.

[6] Tripwire, "Tripwire Vulnerability Scoring System," https://dsimg.ubm-us.net/envelope/160343/293772/1396040281_Tripwire_Vulnerability_Scoring_System_white_paper.pdf, p. 8, 2016. [Online].

[7] C. Alberts, A. Dorofee, J. Stevens, and C. Woody, "Introduction to the OCTAVE approach," Carnegie Mellon University, Tech. Rep., 2003.

[8] N. I. of Standards and Technology, "National Vulnerability Database," https://nvd.nist.gov/, March 2021.

[9] FIRST, "Common Vulnerability Scoring System Version 3.1," FIRST (Forum of Incident Response and Security Teams), Tech. Rep., June 2019. [Online]. Available: https://www.first.org/cvss/v3-1/cvss-v31-specification_r1.pdf

[10] Microsoft, "The STRIDE Threat Model," https://msdn.microsoft.com/en-us/library/ee823878(vcs.20).aspx, pp. 1–3, 2002.

[11] EC-Council, "DREAD threat modeling: An introduction to qualitative and quantitative risk analysis," https://blog.eccouncil.org/dread-threat-modeling-an-introduction-to-qualitative-and-quantitative-risk-analysis/, Dec 2020.

[12] T. Alpcan and T. Basar, *Network security: A decision and game-theoretic approach*. Cambridge University Press, 2010.

[13] C. Pfleeger, *Security in computing*. Prentice Hall, 2015.

 ISBN: 978-1-61208-862-4