ISCTE IUL
REPOSITÓRIO
INSTITUTO UNIVERSITÁRIO DE LISBOA

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:
2026-07-21

Deposited version:
Accepted Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:
Zhao, T., Ongu, D., Gasiba, T., Lechner, U. & Pinto-Albuquerque, M. (2024). A Deep dive into CATS evaluator algorithm: Quantification of the probability in serious game cloud security defense scenarios. In Andreas Bollin, Ivana Bosni, Jennifer Brings, Marian Daun, and Meenakshi Manjunath (Ed.), 2024 36th International Conference on Software Engineering Education and Training (CSEE&T), Proceedings. (pp. 1-5). Würzburg, Germany: IEEE.

Further information on publisher's website:
10.1109/CSEET62301.2024.10663050

Publisher's copyright statement:
This is the peer reviewed version of the following article: Zhao, T., Ongu, D., Gasiba, T., Lechner, U. & Pinto-Albuquerque, M. (2024). A Deep dive into CATS evaluator algorithm: Quantification of the probability in serious game cloud security defense scenarios. In Andreas Bollin, Ivana Bosni, Jennifer Brings, Marian Daun, and Meenakshi Manjunath (Ed.), 2024 36th International Conference on Software Engineering Education and Training (CSEE&T), Proceedings. (pp. 1-5). Würzburg, Germany: IEEE., which has been published in final form at https://dx.doi.org/10.1109/CSEET62301.2024.10663050. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0
The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Serviços de Informação e Documentação, Instituto Universitário de Lisboa (ISCTE-IUL)
Av. das Forças Armadas, Edifício II, 1649-026 Lisboa Portugal
Phone: +(351) 217 903 024 | e-mail: administrador.repositorio@iscte-iul.pt
https://repositorio.iscte-iul.pt

A Deep Dive Into CATS Evaluator Algorithm: Quantification Of The Probability in Serious Game Cloud Security Defense Scenarios

Tiange Zhao
and Didem Ongu
and Tiago Gasiba
Siemens AG
Otto-Hahn-Ring 6
81739, München, Germany
Email: tiange.zhao@siemens.com
Email: didem.oengue@siemens.com
Email: tiago.gasiba@siemens.com

Ulrike Lechner
University of the Bundeswehr Munich
Werner-Heisenberg-Weg 39
85579, Neubiberg, Germany
Email: ulrike.lechner@unibw.de

Maria Pinto-Albuquerque
Instituto Universitário de Lisboa
(ISCTE-IUL)
ISTAR, Av. das Forças Armadas
1649–026 Lisboa, Portugal
Email: maria.albuquerque@iscte-iul.pt

***Abstract*—Cloud deployment has become increasingly common due to its flexibility and business value. However, cloud assets face cybersecurity challenges and need to be configured securely. Industry practitioners must be trained to understand key concepts in cloud security, including 'defense & attack' and 'roles & responsibilities.' A serious game provides an engaging and helpful way to convey such messages. This work introduces the core evaluator algorithm developed for Cloud of Asset and Threats (CATS), a serious game designed to enhance cloud security awareness.**

This work builds upon our previous efforts, focusing on refining the Evaluator algorithm that quantifies the probabilities of the defender strategies as defined by the players to prevent a given attack vector from being successful. We present the results collected from industry training events where the refined algorithm was incorporated into CATS and compare them to the initial implementation. The promising results indicate that with the refinement of the evaluator algorithm, more elements derived from reality are addressed, and the game maintains a similar difficulty level for participants.

I. Introduction

Cloud deployment has become a mature business with increasing connectivity and efficiency. Applying cloud services reduces operators' maintenance workload and provides access to professional management also of the security of applications.

This paper presents the refinement of the Cloud Asset and Threats (CATS) serious game. CATS is a tabletop game played in training in the industry to enhance cloud security. Specifically, our focus is to improve the evaluator algorithm applied in the simulation to calculate the probability of a defense strategy created by the participants successfully blocking a certain attack approach. The plausibility and transparency of the evaluator are important as we would like to simulate real-world cloud security issues. A systematic approach to creating the evaluator is necessary to introduce additional threat scenarios and adapt the game to new threats and technological advances.

This article extends previous work on the design of the CATS game [1], [2], [3] and presents evaluations of the game in industrial training. It is essential to note that the participants in our study were industry practitioners in various industry departments, including IoT devices, healthcare, and energy.

The paper is structured as follows: Section II presents important related work as a foundation of our study. Section III outlines our study methodology. Section IV does the deep dive into CATS background and the evaluator algorithm. Section V illustrates the evaluation process and results. Finally, Section VI concludes our work with a discussion of next steps.

II. Related work

In this section, we present the related work for our research, which consists of IT security awareness, serious games in cybersecurity, and cloud security as a foundation of this paper.

A. IT security awareness

The work of Hänsch et al. on IT security awareness [4] defines three dimensions of IT security awareness: perception, protection, and behavior. Perception is related to knowing about IT security, protection is related to knowing how to protect IT assets, and behavior is related to the intention to actively protect the IT assets.

B. Serious games in cybersecurity

Serious games are games with purposes other than pure entertainment as defined in the field of serious game design by Dörner et al. [5]. Our work follows and extends this baseline.

There are various serious games that aim to raise cybersecurity awareness. Shostack lists security tabletop games on his website [6]. An industry example is Riskio by Hart et al. [7], designed for non-technical individuals in organizations, effectively increasing cybersecurity awareness. Unlike Riskio,

CATS targets industrial practitioners focusing on cloud security concepts.

Another Week at the Office (AWATO) [8] by Ferro et al., based on a systematic literature review, focuses on the human factor and phishing attacks. The evaluation shows it is an effective tool for improving users' cybersecurity awareness. This work supports the idea that serious games can address the awareness issue. However, CATS stands out as one of the first serious games dedicated to cloud security, employing an automatic evaluator algorithm to simulate defense and attacks strategically, providing a reliable probability result.

C. Cloud security

The Cloud Controls Matrix (CCM) [9] from the Cloud Security Alliance (CSA) provides a well-established mapping and comparison of various security controls and cross-checks 44 relevant industrial standards. Notable among these are ISO 27017 and 27018 [10], [11], which specifically address information security controls for cloud services. Apart from the mentioned standards, the MITRE ATT&CK Cloud Matrix categorizes possible attacks against cloud systems based on real-world hacking activities [12]. For each potential attack, effective defense mechanisms are listed. These standards mandate the implementation of a secure development process and provide specific security requirements for cloud security in the industry. This useful information is conveyed to the practitioners in the industry through training where serious games are helpful enhancing methods for raising awareness about cybersecurity issues and solutions.

In our work, the game elements are based on the mapping proposed by the MITRE ATT&CK Cloud Matrix. Additionally, we consider different roles and responsibilities in defending cloud assets. The attack scenarios are abstracted from real-world hacking activities against known cloud assets.

III. Methodology

This work follows Hevner et al.'s design science paradigm [13], focusing on the Design & Implement and Justify & Evaluate cycle.

In the current design iteration, the focus is on the refinement of the evaluator algorithm and systematic mappings with a common vulnerability scoring system (CVSS), and common vulnerabilities and exposures (CVE) that occurred in real-world cloud security incidents. This design iteration is strategically chosen to address identified limitations and continuously improve the serious game. The goal in this iteration is to enhance the realism of attack scenarios and improve the evaluator algorithm to provide players with a better understanding of the consequences of their defense strategies without significantly increasing the game difficulty. Our work contributes to the continuous improvement of the serious game and enhances cloud security awareness among industry practitioners.

The refined evaluator algorithm is tested in 2 game events with 24 industrial practitioners. This targeted testing allows for a focused assessment of the improvements. The evaluation aims to verify the feasibility of the refined evaluator algorithm and measure the performance of the players compared to the previous design.

IV. Algorithm design

In this section, we first present background information about the CATS game, where the evaluator algorithm operates as a core element. Then, we provide technical aspects of the algorithm's design details. To give a straightforward impression, in the last part of this section, we go through the algorithm step-by-step with a concrete example.

A. Background information about CATS game

CATS features six attack scenarios based on real-world incidents, taking approximately 60 minutes to complete. Each scenario follows a three-step kill chain: Initial Access (2 actions), Launch Attack (3 actions), and Make an Impact (1 action). The the current CATS version, the players can utilize the available defense cards and arrange a defense plan to defend the could attack against a given attack scenario.

To defend against a specific attack, players must choose six defense action cards from 23 available cards and assign them to the correct responsibilities. CATS has two responsibilities: business and technical. The former represents business decisions needed for implementing a certain defense (e.g., determining account use policies), while the latter signifies technical implementation tasks (e.g., configuring network segmentation). If a selected card is assigned to the wrong responsibility, it is discarded during the evaluation process and does not contribute to the success of the defense plan. Players can assign a maximum of two cards to business responsibility and four to technical responsibility to strategize the defense plan.

The defense plan and the attack scenario serve as inputs to the evaluator algorithm, producing a probability indicating the likelihood of the defense withstanding the attack. For each attack scenario, a threshold is assigned. If the algorithm's output is higher than the threshold, with all defense cards correctly placed and all attack actions mitigated by at least one defense card, the player successfully solves the scenario and can progress to the next one. If the output probability is lower than the threshold or other conditions are not met, the evaluator proposes hints, allowing the player to adjust the defense plan. There is no limit to the number of trials, and players are encouraged to experiment with different combinations and learn about cloud security.

In this work, we focus on a refinement of the mapping of defense and attack actions by introducing the CVSS scoring system into the conditional probability calculation.

B. Impact of the defense actions

There is a many-to-many mapping between defense and attack actions. When a defense action mitigates a vulnerability, the defense action has an impact on its CVSS vector.

In CATS, the base metric group of CVSS vectors is used. We apply CVSS 3.1 [14], as it is the mainstream version of CVSS for the time being.

A valid defense has two different types of impact upon an attack. It could either reduce the factor by one value or eliminate the potential impact on a certain vector of the system or component.

Let's assume there is an attack action Network Service Discovery [15] mapped to the CVE cve-2020-1206 [16]. The vector in the base metric of cve-2020-1206 is:

CVSS:3.1/AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:N/A:N

with a score of 7.5 (High). We apply the defense action Network Segmentation, the impacts of which are to reduce attack vector (AV) and scope (S) by one value. AV is updated from Network (N) to Adjacent (A). Since the Scope of the original CVE is Unchanged (U), which is already the lowest value and cannot be further reduced, the second impact does not have any substantial influence on the given attack action. By applying Network Segmentation, the vulnerability's vector is updated to:

CVSS:3.1/AV:A/AC:L/PR:N/UI:N/S:U/C:H/I:N/A:N

with a score of 6.5 (Medium).

C. Mapping of attacks and common vulnerabilities and exposures (CVE)

The game CATS requires players to address six attack scenarios. Each attack scenario simulates a step-by-step attack consisting of six attack actions. The attack actions are possible due to the vulnerabilities that might exist in cloud assets or are due to misconfigurations. In this work, we endeavored to map each attack action to an existing vulnerability in the first place. When there were no suitable ones to take, we derived the CVSS Vector of our own.

The CVE mapping process involved searching with keywords related to basic attacks from MITRE ATT&CK database, identifying several CVE records, and selecting the one with the highest vulnerability score to apply the worst-case scenario in the game. It's important to note that the game serves as a simulation to provide an understanding of fundamental concepts. While the mapping process is thorough, it doesn't cover all possible real-world scenarios.

For instance, brute Force is an attack technique that adversaries may use to gain access to accounts when passwords are unknown or when password hashes are obtained. Without knowing the password for an account or set of accounts, an adversary may systematically guess the password using a repetitive or iterative mechanism [17]. In our game, we mapped this attack action to CVE-2020-14494. This vulnerability is due to a flaw in an authentication mechanism that does not provide sufficient complexity to protect against brute force attacks, which may allow unauthorized users to access the system after no more than a fixed maximum number of attempts [18]. CVE-2020-14494 is a critical vulnerability, and the CVSS vector is CVSS:3.1/AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:H/A:H with a score of 9.8 (Critical) out of 10.0. For all the mapped vulnerabilities, we applied the CVSS version 3.1.

D. Mapping of attack and defense actions

Multiple defense actions can defend one attack action, and one defense action could also mitigate multiple attack actions. This many-to-many mapping is illustrated in MITRE ATT&CK cloud matrix [12]. For instance, the defense "Application Developer Guidance" refers to the defense that developers should be trained to develop applications securely. This activity is helpful in preventing the attack "Abuse credential" when the debugging and development access to the application cannot be misused.

E. Algorithm walk-through of an example

In this section, we walk through the algorithm with a example: the attacker applies "Abuse Credential" and "Cloud Infrastructure Discovery" in step 1 to gain access. In step 2, the attacker uses the attack actions (AA) "Abuse Trusted Relationship," "Monitoring Escaping," and "Impair Defense" to launch the attack. In step 3, the attacker conducts "Resource Hijacking" to make an impact. This is shown in the right part of Figure 1.

Within one step, attack actions (AAs) are considered in parallel, which means the attack success rate for this step will only be reduced to 0% if all attacks within this step are mitigated to 0%. If there is one undefended attack with a 100% success rate, the attack success rate would be 100%. This case symbolizes that the attacker finds an open vulnerability and applies an exploit; therefore, the attacker can bypass this step without any trouble. We assume the normalized percentage of an attack m in step n is p_{nm}, then the attack success rate of step n, p_n, is:

$$p_n = 1 - \prod_{1...m}(1 - p_{nm}).$$

The defense success rate of step n is:

$$\prod_{1...m}(1 - p_{nm}).$$

Then we consider the three steps in serial order. If an attacker attacks the cloud asset in a three-step approach, he or she must succeed in all the steps to finally compromise the cloud asset. Therefore, if we assume the final attack success rate is p, then:

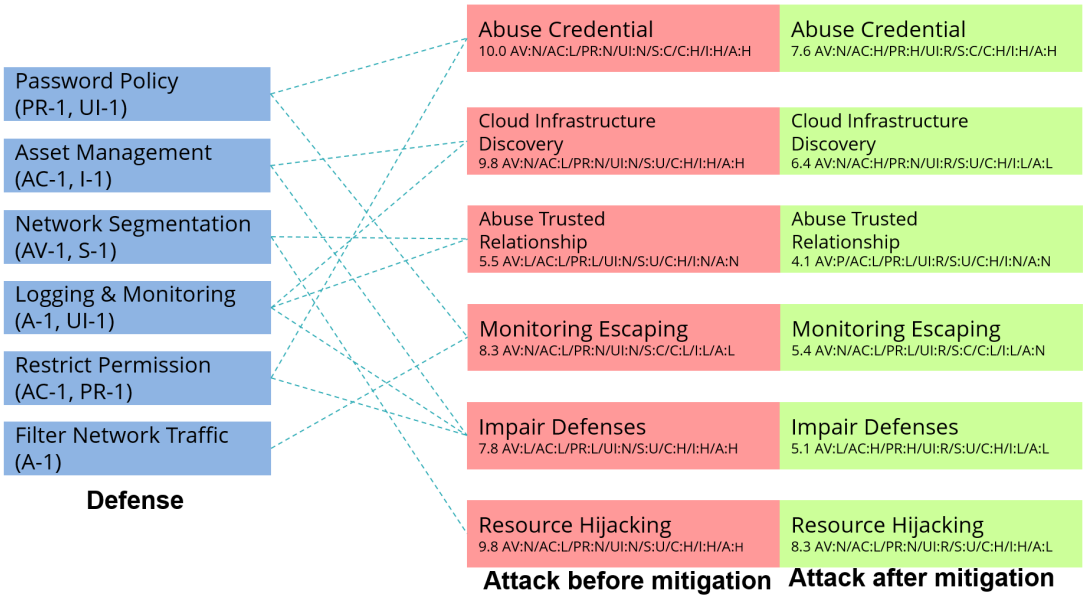
$$p = \prod_{1...3}(p_n).$$

The final defense success rate p_{final} can be derived:

$$p_{final} = 1 - p.$$

Following this logic, we can derive the final defense success rate of 3% ($p_1 = 100\%, p_2 = 98\%, p_3 = 98\%$). The defense actions are shown in the left part in Figure 1 marked as "Defense". We assume that those cards are assigned to the correct role, therefore the impact remains valid. After applying the defense actions (DA), as shown on the right of Figure 1, the CVSS base score of each attack action (AA) drops. After applying the calculation on the new values, we

Fig. 1. Mapped vulnerability before mitigation.



derive that the defending probability has increased from 3% (before mitigation) to 34% (after mitigation, ($p_1 = 91\%, p_2 = 86\%, p_3 = 83\%$)). This is still not good enough to solve the scenario, but there will be hints proposed to help the player improve the quality of the defense plan and finally achieve the required threshold.

Note that the players need to assign the activities to the correct roles in the game. If a defense card and activity are incorrectly assigned, it is ineffective in mitigating the attack and its contribution to the defense is "zero". Our walkthrough assumes that all cards are assigned to the proper roles.

V. Evaluation

In this section, we present the method we use to evaluate the implemented "Evaluator" algorithm. Then, we show the collected results and discuss them.

A. Evaluation Method

We evaluate the new algorithm by applying it to in-game events and comparing the statistics of the new algorithm with the previous version of CATS. At the end of each game event, we also invited the participants to an open discussion to exchange ideas and collect feedback in the form of semi-structured interviews (SSI). We organized ten game events before refining the evaluator algorithm and two with the new Evaluator algorithm. For the two-game events in the current design iteration, 24 industrial practitioners participated as players. The game event took place on 2023-03-22 (14 players) and 2023-04-19 (10 players). In all of the 12 game events, the players are practitioners in the industry with similar experience levels. Further demographic information was not collected since it was irrelevant to the data analysis.

B. Evaluation Results

We present in Table I the comparison of the "historic" and the new Evaluator algorithm. The evaluation results were obtained through in-game data collection and analysis.

The in-game data includes details of each player's actions, submissions, and outcomes during the gameplay. This granular data allowed us to conduct a detailed analysis of player performance, submission patterns, and defense strategies' effectiveness. The in-game data was processed and analyzed to generate the quantitative metrics presented in Table I. The data of the two-game events are combined with a total sample number of 24.

Comparing the players' performance before (sample number 123) and after the algorithm refinement (sample number 24), we observed that with the new algorithm, the players took fewer submissions per scenario to solve the scenario. The average time the players spent on each scenario drops slightly in the new algorithm, On average, the players tend to make more improvement in terms of increasing defense success rate in each submission in the new algorithm. However, in the game events with the new algorithm, players seem to make more mistakes in the role assignment per scenario.

C. Discussion

Participants enjoyed the CATS game and found it beneficial for enhancing their cloud security knowledge. This remained unchanged for the ten historical game events and the two new game events. The use of CVSS in CATS allows for simulating attacks and reflecting more elements from reality. All metrics combined from table I, we could conclude that the algorithm refinement with more elements reflecting reality in the current design iteration did not significantly increase the game's difficulty.

TABLE I
GAME EVENT STATISTIC: COMPARING THE IMPROVED ALGORITHM TO THE ORIGINAL ALGORITHM

	Original Algorithm	**New Algorithm**
Average no. of submissions per scenario	7.85	7.30
Average time to solve each scenario	5 min. 45 sec.	5 min. 18 sec.
Average improvement in each submission	5.21	6.55
No. of mistakes in role assignment per scenario	2.11	3.04

Post-game discussions through SSIs offered qualitative insights, addressing challenges and suggesting improvements. A comparison between the original and new algorithms revealed consistent participant engagement without increased difficulty, ensuring a steady level of challenge.

VI. CONCLUSION AND FUTURE WORK

In our work, we improved the core evaluator algorithm with CVSS and MITRE ATT&CK. By involving the mentioned elements above, we managed to:

1) Consider the different impacts of each defense action individually.

2) Enhance the real-world characteristics of the simulated CATS game without significantly increasing the difficulty.

The study is conducted under the guidance of the design science paradigm, which is an approach that combines scientific research with practical design to create innovative solutions to real-world problems. As for the evaluation, we organized two game events with industry practitioners. The collected feedback shows that the players find the game enjoyable and helpful in raising awareness about the challenges in cloud security and exercising the skills for designing a successful defense strategy.

The goal of this design iteration was to enhance the realism of attack scenarios and improve the evaluator algorithm to provide players with a better understanding of the consequences of their defense strategies. The fact that players reacted similarly suggests that the difficulty level was effectively maintained, preventing an unintended increase in complexity. This outcome aligns with our objective to refine the algorithm without imposing unnecessary challenges on the players.

In our future work, we would like to conduct more game events and invite cybersecurity experts to review the mapping and the algorithm workflow. Based on the collected feedback, further improvement and adjustment can be carried out.

ACKNOWLEDGMENT

This research task was partially supported by Fundação para a Ciência e a Tecnologia, I.P. (FCT) [ISTAR Projects: UIDB/04466/2020 and UIDP/04466/2020]. Ulrike Lechner acknowledges funding for project CONTAIN by the Bundesministerium für Bildung und Forschung (FKZ 13N16581). Tiange Zhao and Tiago Gasiba acknowledge the funding provided by the Bundesministerium für Bildung und Forschung (BMBF) for the project CONTAIN with the number 13N16585.

REFERENCES

[1] T. Zhao, U. Lechner, M. Pinto-Albuquerque, E. Ata, and T. Gasiba, "Cats: A serious game in industry towards stronger cloud security," in *Ubiquitous Security*, G. Wang, K.-K. R. Choo, J. Wu, and E. Damiani, Eds. Singapore: Springer Nature Singapore, 2023, pp. 64–82.

[2] T. Zhao, U. Lechner, M. Pinto-Albuquerque, and E. Ata, "Cloud of Assets and Threats: A Playful Method to Raise Awareness for Cloud Security in Industry," in *Third International Computer Programming Education Conference (ICPEC 2022)*, ser. Open Access Series in Informatics (OASIcs), A. Simões and J. a. C. Silva, Eds., vol. 102. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, pp. 6:1–6:13. [Online]. Available: https://drops.dagstuhl.de/opus/volltexte/2022/16610

[3] T. Zhao, U. Lechner, M. Pinto-Albuquerque, and D. Ongu, "An ontology-based model for evaluating cloud attack scenarios in cats – a serious game in cloud security," in *2023 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 2023, pp. 1–9.

[4] N. Hänsch and Z. Benenson, "Specifying IT security awareness," in *25th International Workshop on Database and Expert Systems Applications*, IEEE. IEEE headquater address3 Park Avenue, 17th Floor New York, NY 10016-5997 USA: IEEE, 2014, pp. 326–330.

[5] R. Dörner, S. Göbel, W. Effelsberg, and J. Wiemeyer, *Serious Games: Foundations, Concepts and Practice*. Springer Nature One New York Plaza, Suite 4600 New York NY 10004-1562 USA: Springer, 2016.

[6] A. Shostack, "Tabletop security games & cards," https://https://shostack.org/games.html, 2021.

[7] S. Hart, A. Margheri, F. Paci, and V. Sassone, "Riskio: A serious game for cyber security awareness and education," *Computers & Security*, vol. 95, p. 101827, 2020.

[8] L. S. Ferro, A. Marrella, T. Catarci, F. Sapio, A. Parenti, and M. De Santis, "AWATO: A Serious Game to Improve Cybersecurity Awareness," in *HCI in Games*, X. Fang, Ed. Cham: Springer International Publishing, 2022, pp. 508–529.

[9] C. C. M. W. Group, "Cloud security alliance: Cloud controls matrix v4," https://cloudsecurityalliance.org/artifacts/cloud-controls-matrix-v4/, 2021.

[10] ISO27017, "ISO/IEC 27017:2015 Information technology — Security techniques — Code of practice for information security controls based on ISO/IEC 27002 for cloud services," https://www.iso.org/standard/43757.html, 2015.

[11] ISO27018, "ISO/IEC 27018:2019Information technology — Security techniques — Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors," https://www.iso.org/standard/76559.html, 2019.

[12] T. M. Corporation, "MITRE ATT&CK cloud matrix," https://attack.mitre.org/versions/v8/matrices/enterprise/cloud/, 2020.

[13] A. Hevner, "A three cycle view of design science research," *Scandinavian Journal of Information Systems*, vol. 19, pp. 1–6, 2007. [Online]. Available: http://aisel.aisnet.org/sjis/vol19/iss2/4

[14] N. Organization, "Vulnerability metrics," 2023. [Online]. Available: https://nvd.nist.gov/vuln-metrics/cvss

[15] MITRE, "Network Service Discovery," https://attack.mitre.org/versions/v13/techniques/T1046/, 2017.

[16] ——, "CVE-2020-1206 Detail," https://nvd.nist.gov/vuln/detail/cve-2020-1206, 2020.

[17] ——, "Brute Force," https://attack.mitre.org/versions/v13/techniques/T1110/, 2017.

[18] ICS-CERT, "CVE-2020-14494 Detail," https://nvd.nist.gov/vuln/detail/CVE-2020-14494, 2020.