From Prey to Predator: A Use Case for Using Active Defense to Reshape the Asymmetrical Balance in Cyber Defense

Pei-Yu Huang
McKelvey School of Engineering
Washington University in St. Louis
St. Louis, Missouri
h.peiyu@wustl.edu

Yi-Ting Huang
Dept. Electrical Engineering
National Taiwan University of Science and Technology
Taipei, Taiwan
ythuang@mail.ntust.edu.tw

Yeali S. Sun
Dept. Information Management
National Taiwan University
Taipei, Taiwan
sunny@ntu.edu.tw

Meng Chang Chen
Research Center for Information Technology Innovation
Academia Sinica
Taipei, Taiwan
mcc@iis.sinica.edu.tw

***Abstract*—Cyber-security countermeasures predominantly rely on a passive approach of waiting for adversaries to trigger predefined rules. However, active defense involves actively engaging with adversaries to observe, affect, and elicit attack behaviors by providing misleading information. This approach transforms the asymmetric nature of cyber-security defense. We conducted simulations and analyzed the techniques used in a real-world cyber-attack based on the MITRE ATT&CK framework. Then, based on the MITRE Engage framework, we identified potential use cases for implementing active defense as a countermeasure.**

***Index Terms*—active defense, deception, TTPs, proactive defense**

1. Introduction

The surge of cyber-attacks in recent decades has highlighted the crucial role of cyber-security across all industries and organizations. Advanced Persistent Threat (APT), a prolonged and illicit activity that targets victims' systems with malicious actions, has presented significant challenges for companies and organizations to manage. The asymmetrical nature of cyber security creates a significant disadvantage for the defense party on the APT battlefield. A typical APT attack comprises various attack actions, and the actions are spread across multiple system logs, making it computationally challenging for defenders to observe malicious behaviors. Moreover, adversaries frequently evade pre-defined detection rules and exploit zero-day vulnerabilities, exacerbating the situation for defenders. In short, APT attacks are becoming increasingly sophisticated and challenging to detect, presenting a significant threat to organizations' cyber security.

By engaging with adversaries, active defense is a proactive approach that can transform the asymmetrical nature of cyber defense. ”With traditional cyber defense, the adversary only needs to be right once, but with adversary engagement, the adversary only needs to be wrong once [1].” This paradigm shift in the cyber-security industry provides defenders with a first-mover advantage, giving them an edge over attackers.

This article explores the concept of active defense and provides insights into how and where it can be implemented based on a real-world attack. We also highlight some of the key challenges that defenders must be aware of and explore future possibilities in this emerging field of cyber security.

2. Background and Related Work

2.1. Active Defense

Active defense, also referred to as “proactive defense” is aiming to negatively affect adversaries' attack operations. It aims to increase the adversary's operation cost and efforts, decrease the loss affected by their attacks, expose their vulnerabilities, and elicit more attack Tactics, Techniques, and Procedures (TTP) [2]. This can be achieved primarily through two approaches: Denial and Deception. The classical denial and deception matrix [3][4] shown in Table 1 considered both denial and deception to be two essential methods in deception, which focused on concealing and revealing information respectively.

We treat these two terms separately. Denial approach can be performed without using deception, while all deception involves some form of denial [5]. To prevent confusion, we used the following definitions:

Denial: To hinder adversaries from collecting valuable information and to weaken their operational effectiveness and efforts. For instance, manipulating system networks to limit and throttle network speeds, and blocking network connections to slow down attacks and impair adversary actions.

Deception: To mislead adversaries by creating fictitious system artifacts or stories combined with a portion of facts to perplex their perception of the system network.

TABLE 1. D&D METHODS MATRIX. ADAPTED FROM [3][4].

Deception Objects	Revealing	Concealing
Facts	Show the real	Hiding the real (dissimulation)
Fictions	Showing the false (simulation)	Hiding the false

For instance, manipulating information increases adversaries' uncertainty and ambiguity of the system environment by concealing and revealing fake and real information.

Traditional defense methods such as Intrusion Detection Systems (IDS) and Security Information and Event Management (SIEM) are behavior-based and anomaly-based detections that are dealing with big data burdens. Active defense, on the other hand, is an action-based detection that deals with a ”right data” problem. It detects malicious activities by observing and interacting with adversaries. Through the use of decoys and traps that adversaries are not meant to touch, active defense can quickly and accurately identify abnormal interactions, resulting in a lower false-positive (FP) than traditional IDS/SIEM methods. Security product vendors have been focusing on reducing false negatives. However, reducing FP is equally important [6]. The costs and time to identify false alarms (i.e., benign activities but alarmed) and benign triggers (i.e., malicious activities truly alarmed) can be substantial, causing dissatisfaction among defenders [7].

Active defense employing deceptive methods not only slows down adversarial activities but also creates self-doubt [8][9], affecting their strategies and cognitive reactions. Even if adversaries try other techniques, they would end up wasting more time and resources.

2.2. MITRE ATT&CK®

MITRE ATT&CK framework documents TTP [10]. Tactic refers to the underlying objectives or purpose behind adversarial actions. Each action an APT creates can be subjected to different tactics. Technique refers to the action performed to achieve a tactic. Each adversarial action or series of actions has tactics, and each tactic can be accomplished with various techniques. Procedures are concise implementations of techniques by specific APT groups or software. Recent studies [11] [12] [13] leverage ATT&CK as a rich source of adversary knowledge to develop their defending systems. The framework supports enterprises to dissect an APT attack and understand why and how it operated to harm their business.

2.3. MITRE Engage™

In terms of active defense, MITRE Engage framework was introduced in April 2022 [1]. It is a knowledge base of active defense, which assists defenders in understanding approaches to counter adversaries. The Engage framework consists of five goals, nine approaches, and thirty-one activities.

- **Goal:** The five goals are Prepare, Expose, Affect, Elicit, and Understand. Prepare and Understand are the inputs and outputs of active defense, which determine components to strategize and how to revise after each attack. Expose is to detect and send alerts when adversaries are in the environment. Affect is to negatively impact adversaries' actions and operations. Elicit is to lure or trick adversaries to perform more TTPs.
- **Approach:** Approaches correspond to the tactics in ATT&CK. Approaches represent the objectives of certain active defense activities and utilize the concept of denial and deception.
- **Activity:** Activities correspond to the techniques in ATT&CK. The thirty-one activities are driven by real adversarial actions and represent how approaches are implemented.

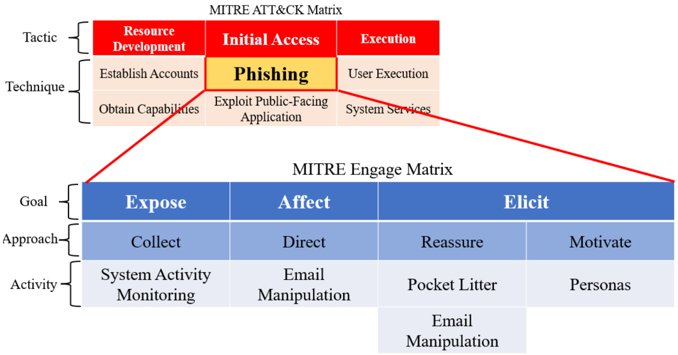

Figure 1. Phishing technique mapped to Engage Matrix, showing five activities with different goals and approaches are available.

ATT&CK is able to map to the Engage framework (Fig. 1). Defenders can utilize ATT&CK supplement with Engage framework to understand attack scenarios and active defense implementations. The Engage framework enriches the definition of active defense. It keeps Denial and Deception as its core ideas, using them to impact adversaries' behaviors.

3. Use Case: Deploying Active Defense

In May 2020, the Chinese Petroleum Corporation (CPC) in Taiwan, a natural gas corporation, fell victim to a ransomware attack that caused a system outage. This disrupted customers' ability to pay through CPC's VIP cards and electronic transaction applications [14]. This attack is a typical example of a targeted attack on Active Directory (AD), which can be catastrophic for organizations' operations if compromised. We simulated this case (Fig. 2) to analyze the tactics and techniques employed based on ATT&CK, and we identify possible active defense implementations based on the Engage activities.

Initial Access. The attack on CPC's network originated from a compromised internal PC or web server (marked as (1) in Fig. 2) with a dwell time of around 1400 days [14]. Adversaries likely used techniques such as T1190: Exploit Public-Facing Application and T1566: Phishing to gain initial access. Attacks like T1190 focus on exploiting vulnerabilities in public-facing applications, which are services available to the public and also have access to the internal network. Adversaries may try to sniff packets,

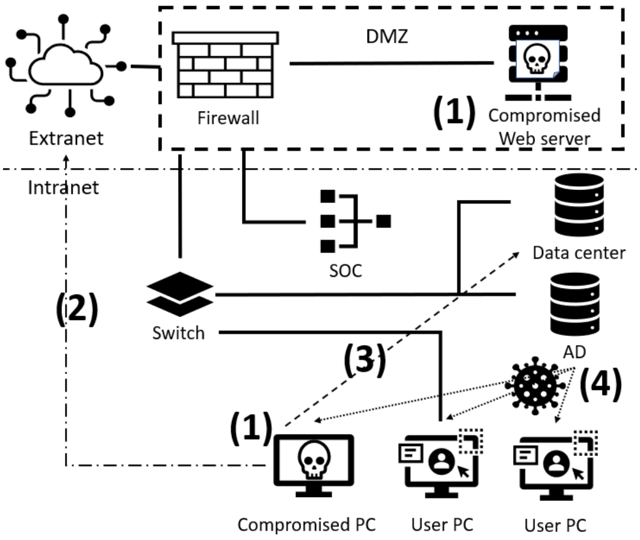

Figure 2. Simulated Attack Process

scan for open ports, and inject malicious code or scripts into the applications to execute unwanted commands. In this case, adversaries could have utilized this technique to gain access to the web server and establish an initial foothold in the system. In addition, T1566 involves using social engineering tactics to gain initial access to a target's system. This includes sending fraudulent emails containing malicious links or attachments that, once clicked, deploy malware onto the system, giving the adversary an initial foothold in the network.

Command and Control (C2). Once compromising an endpoint, adversaries established a connection back to themselves from the compromised device like (2) in Fig. 2 using T1021: Remote Services and T1071: Application Layer Protocol. C2 connection is stealthy and tries to evade detection. Adversaries can blend in with normal traffic created by the application layer protocols and transfer data or establish a connection. C2 can be created through Domain Name System (DNS) Tunneling (T1071). Since DNS queries are considered normal and expected traffic to move in and out of the firewall, it embeds data or messages in DNS requests to the C2 server. Then, with the connection to the compromised device, the adversary can plot their next move and advance their attack.

Discovery. When obtaining C2 connection, the adversary would start an internal reconnaissance like (3) in Fig. 2 to determine their next step. This process may involve using techniques such as T1049: System Network Connections Discovery, T1018: Remote System Discovery, T1016.001: System Network Configuration Discovery: Internet Connection Discovery, and T1046: Network Service Discovery. Discovery is a crucial step for adversaries to gather information about the target network environment. Adversaries often use automated tools to scan all ports on every IP address to locate valuable servers. Once adversaries have located valuable servers or assets, they will attempt to move laterally toward the target. A common target is AD. Taking control of the AD allows adversaries to identify users and hosts, understand group policies and permissions, and access core services such as asset management systems, human resource systems, and web server databases.

Credential Access & Collection. Adversaries were able to gain access to AD by posing as legitimate users using T1550: Use Alternate Authentication Material and T1056: Input Capture. T1550 allows adversaries to move laterally in the network by bypassing access controls, often by using a sub-technique called ”Pass the Hash (PtH)” to extract hashed credentials from compromised devices. With PtH, adversaries can access AD as regular users. T1056, on the other hand, allows adversaries to collect user input through techniques like keylogging. By intercepting keystrokes, adversaries can collect passwords, including those of users who have logged on to the compromised device using a DC account. Therefore, adversaries may have intercepted passwords during this attack.

Persistence & Privilege Escalation. After gaining access to AD, the adversaries went on to distribute ransomware, as seen in (4) of Fig. 2, using T1053.003: Scheduled Task/Job. They forged a Group Policy Object (GPO) and schedule a task for all computers to download and execute the ransomware automatically to target a large number of systems at once.

Impact. The ransomware was downloaded and executed on all PCs as scheduled when the employee booted the PC using T1486: Data Encrypted for Impact. The ransomware then proceeded to encrypt files and folders in computers, after which the victim received a message to pay for the decryption key. This step also marked the completion of the APT mission.

4. Proposed Active Defense Implementations

We mapped Engage activities (written in italics) to ATT&CK in order to address the attacks. We suggest actual implementations of these activities as part of an active defense strategy. Each implementation can complement the other and does not require a particular deployment sequence.

Initial Access. Active defense provides an additional layer of protection against phishing attacks by using decoy mailboxes. These decoys are designed to attract adversaries and prompt them to engage with them, providing defenders with early warnings of potential attacks. This approach, coupled with traditional prevention methods such as anti-spoofing mechanisms, authentication rules, and blacklisting, can be an effective way to detect and prevent phishing attacks.

Adversaries use automated tools to harvest email addresses from websites by searching for the symbol “@”. Decoy mailboxes in the format of regular emails can be easily extracted through *Email Manipulation* techniques. These decoys are placed on orphan pages of the website to ensure they can be harvested by bots but not by regular customers. It is important to note that the email addresses of general employees are still visible to adversaries and customers, whereas only the decoy emails are accessible to adversaries.

Adversaries typically send out large numbers of phishing emails automatically and simultaneously. Decoy mailboxes are not intended to receive any legitimate emails, and any interactions with these decoys are considered unauthorized access. Therefore, when decoy emails receive any messages, defenders can immediately block the associated email addresses and IPs to prevent regular employees from viewing them. Furthermore, by isolating malicious links and objects in a virtual machine (VM), defenders can monitor their behavior and take appropriate action.

Besides creating a deceptive environment, it's important to create a plausible backstory to support the narrative. This is where *Personas* come in, which are fictitious identities used to make the deception more believable. For example, a persona may involve creating a fake employee profile on a social media platform like LinkedIn, complete with a decoy email address. To increase the realism of the persona, a backstory (also known as *Pocket Litter*) may be created to include details such as hobbies, personal and professional interactions, profile data, and updates. All of these elements work together to make the decoy email address appear to have an active and credible owner, tricking adversaries into engaging with the persona.

Moreover, adversaries may leverage scanning tools to identify open ports and exploit vulnerabilities to gain access to web servers. To counter this, defenders can set up a decoy web server that mimics the appearance of a legitimate server, but actually logs and monitors the activities of any potential intruders (marked as (1) in Fig. 3).

In the real world, most organizational applications have varying patch levels. Defenders can deploy multiple decoy web servers with different patch levels, thus ensuring *Application Diversity*. This can increase the realism of the network system and either motivate or demotivate the adversary.

We suggest some other activities that are also relevant to this approach. Through *Network Manipulation*, the decoy web server can interact with the adversary and alert the security team, by guiding the adversary toward the decoys. Defenders can also introduce *Introduced Vulnerabilities* to redirect the adversary from the real server and gather information about their capabilities and resources. These active defense measures allow defenders to receive alerts before the adversary can penetrate the system. They also divert adversary activity towards decoys, providing reaction time for defenders.

Discovery. To thwart adversaries who use port and service scans to probe, defenders can install a module at the endpoints that monitor any scans directed to a closed port, which is a form of *Network Manipulation*. This means that scans sent to a non-existent device or service are deemed suspicious. For example, port 80 and port 443 are typically used for web services, while port 53 is used for DNS servers. If an IP address attempts to access port 443 on a DNS server, an alert is generated to indicate abnormal scans on the network. The module then redirects the query to a decoy server pretending to be a web server, which is actually a form of *Software Manipulation*. Defenders can then observe the suspicious activity and isolate it from the real network. This form of software manipulation wastes adversaries' time as they plan their next move on a decoy and causes confusion. Any scans to decoy endpoints Especially, the scans sent to decoy endpoints (marked as (4) in Fig. 3) would also be considered suspicious.

If the AD receives suspicious scans, defenders can use a decoy AD (marked as (2) in Fig. 3) to redirect the queries to itself, pretending to be the legitimate one. This not only conceals the real AD and production assets from the endpoint but also reduces the attack surface. The decoy AD can then provide fake information, such as IP addresses and domain account names, that appear genuine enough to maintain a realistic deceptive environment. Deploying decoy servers or AD can significantly reduce the adversary's process time and consume their resources interacting with decoys. Additionally, the use of *Pocket Litter*, such as decoy network segments or virtual local area networks (VLANs), can enhance the credibility of the system environment.

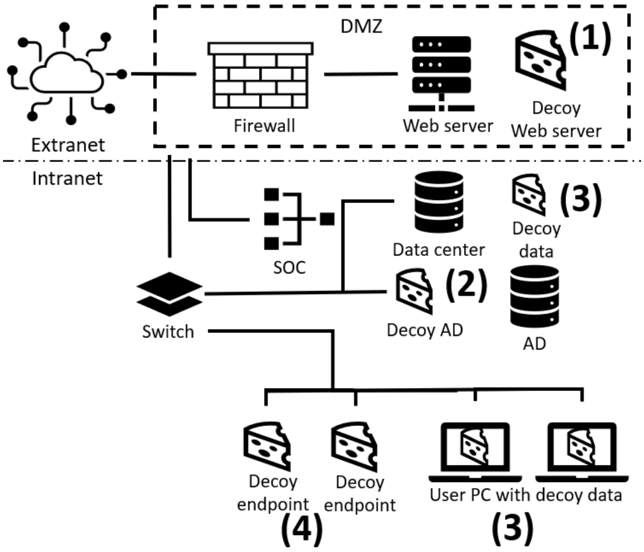

Figure 3. Proposed decoys types

Credential Access & Collection. In the event that adversaries acquire login credentials for the domain controller through a compromised computer, they can gain control of the AD as an authorized user. It is common for adversaries to attempt to obtain information such as usernames, passwords, access keys, and tokens by extracting password hash from either memory or hard disks (e.g., account manager or password/shadow file). Defenders can strategically place fake credentials (marked as (3) in Fig. 3) in these locations, intentionally making them appear vulnerable or challenging to crack (i.e., *Information Manipulation*). If an adversary uses these decoys, alerts are triggered. Once defenders detect suspicious activity in the network, they can leverage *Security Controls*, such as modifying login policies, to make it easier for the adversary to log in to monitored environments. This allows defenders to observe and elicit information on more advanced TTPs employed by the adversary. To increase the effectiveness of decoys, they can take the form of an entire endpoint or be represented as decoy credentials within a legitimate PC, which is known as *Artifact Diversity*. The credibility of decoy endpoints (marked as (3) in Fig. 3) can be improved by providing them with *Pocket Litters* such as browsing history, pictures, installed software, and connection history. Moreover, decoy files can also be mixed with non-essential real files and encrypted using base64 or Advanced Encryption Standard (AES) to enhance their credibility.

This kind of diversity presents additional attack surfaces that can motivate and demotivate adversaries at the same time. For each step, they need to verify whether the credentials are legitimate and simultaneously avoid possible traps with exceptional caution.

Impact. This is the scenario in which ransomware has been executed successfully and has begun encrypting data. It is the last resort for defenders to detect this activity if none of the active defense measures were deployed during the earlier stages or have failed to function properly. To increase the chances of detecting ransomware activity, defenders can display various system artifacts, such as browser cookies, directories, and files, on the network to create *Artifact Diversity*, and all of these artifacts are closely monitored for any interactions. If any decoy files (marked as (3) in Fig. 3) are modified, this indicates that ransomware has been executed, and defenders are alerted. Nonessential files are also put among decoy files to minimize overall impact.

Concurrently, the ransomware is redirected to a VM and isolated from the system network (i.e., *Software Manipulation*. The VM would contain decoy files, such as .doc, .xlsx, .pdf, and .mp4, are included, exceeding the quantity of actual files by one hundred times. This serves to slow down the ransomware's encryption rate and create the appearance that it is still carrying out its intended function. (i.e., *Information Manipulation*).

In summary, active defense employs a range of tactics to expose the actions of adversaries and encourage them to persist with their attacks. Although the above methods may seem disparate, contemporary active defense solutions [15] can create decoy devices from a centralized control system. Defenders can utilize customized or automated services from product vendors or create their own decoy devices.

5. Possibilities and Challenges

We discuss the possibilities and challenges of applying active defense to enterprises.

5.1. Possibilities

Evaluating the effectiveness of active defense remains a challenge. Having a standardized benchmark would be a valuable way to measure the performance of a defender's security team. Such a benchmark would provide insight into the number of attacks attempts that the security team has prevented and the additional TTPs that they have collected. MITRE Engenuity ATT&CK Deception Evaluation [16] assessed the effectiveness of active defense and deception. The evaluation revealed the TTPs identified by participants and how decoys were displayed to adversaries, but it did not show how adversaries interacted with the products or how the products could elicit additional TTPs from them.

Active defense still lacks a standardized way to measure active defense products or performance equally. Future possibilities include establishing a standardized measurement, which would provide defenders with a more comprehensive understanding of the number of attack attempts made, how long they have persisted, and how the active defense strategy has evolved over time.

5.2. Blind spots

Maintaining active defense decoys everywhere is not feasible, which creates blind spots in areas without coverage. Having too much coverage incurs high maintenance costs while having too little coverage results in blind spots for detection. Thus, balancing the coverage and budget is crucial. Additionally, since active defense has limited coverage, traditional log analysis methods still play a significant role in supplementing this limitation.

5.3. Require expertise to operate

While active defense offers several benefits, it cannot be achieved without the presence of a professional incident response team. The team is crucial for implementing active defense as they are responsible for preparing, handling alerts, and assessing the outcome.

The implementation of active defense involves three key stages: Prepare, Operate, and Understand [1]. Prepare involves defining the expected outcomes, understanding the system environment, determining the desired amount of decoys and their coverage, and setting a gating threshold to interfere when decoy interactions exceed acceptable risks. Without proper preparation, it is difficult to determine which active defense operations work best for the defender's systems. Operate involves exposing, affecting, and eliciting the adversary. A team must react quickly to alerts with predefined strategies, and also be prepared to improvise based on the situation. Understand involves assessing and providing feedback on how the active defense performed in the previous stage. This includes analyzing collected TTPs and cyber intelligence, checking whether implementation aligned with objectives, and revising strategies based on analysis results.

The balance between decoy coverage and budget is crucial, as higher coverage requires hiring a sophisticated incident response team to manage alerts, while lower coverage may make active defense almost invisible and difficult to trigger, resulting in minimal TTP information.

6. Conclusion

Active defense is a rapidly growing sector within the cybersecurity industry, representing a paradigm shift in the fight against cyber attacks. By simulating an APT attack scenario based on CPC's case, we propose various active defense strategies utilizing denial and deception techniques. These tactics can impede adversaries' operations by providing false information that entices them to reveal their intentions, leading to high-quality alerts that increase their resource costs. Moreover, active defense encourages adversaries to use more advanced TTPs, providing defenders with a more in-depth understanding of their opponents' actions. Future research may focus on designing the believability and interaction factors for decoys.

References

[1] *MITRE Engage*. https://engage.mitre.org/. Last accessed 30 July 2022.

[2] *Adversarial Tactics, Techniques & Common Knowledge (ATT&CK), v11*. https://attack.mitre.org. Last accessed 30 July 2022.

[3] Kristin E Heckman et al. "Cyber denial, deception and counter deception". In: *Advances in Information Security* 64 (2015), pp. 15–23.

[4] Michael Bennett and Edward Waltz. *Counterdeception principles and applications for national security*. Artech House, 2007.

[5] Robert M Clark and William L Mitchell. *Deception: Counterdeception and counterintelligence*. CQ Press, 2018.

[6] Chris Crowley and John Pescatore. "Common and best practices for security operations centers: Results of the 2019 SOC survey". In: *SANS, Bethesda, MD, USA, Tech. Rep* (2019).

[7] Bushra A Alahmadi, Louise Axon, and Ivan Martinovic. "99% False Positives: A Qualitative Study of SOC Analysts' Perspectives on Security Alarms". In: *Proceedings of the 31st USENIX Security Symposium (USENIX Security), Boston, MA, USA*. 2022, pp. 10–12.

[8] Kimberly J Ferguson-Walter, Dana S LaFon, and TB Shade. "Friend or faux: deception for cyber defense". In: *Journal of Information Warfare* 16.2 (2017), pp. 28–42.

[9] Kimberly Ferguson-Walter et al. *The Tularosa Study: An Experimental Design and Implementation to Quantify the Effectiveness of Cyber Deception.* Tech. rep. Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2018.

[10] Blake E Strom et al. *MITRE ATT&CK: Design and philosophy*. https://www.mitre.org/sites/default/files/publications/pr-18-0944-11-mitre-attack-design-and-philosophy.pdf. 2018.

[11] Valentine Legoy et al. *Automated Retrieval of ATT&CK Tactics and Techniques for Cyber Threat Reports*. 2020. arXiv: 2004.14322 [cs.CR].

[12] Rawan Al-Shaer, Jonathan M. Spring, and Eliana Christou. "Learning the Associations of MITRE ATT&CK Adversarial Techniques". In: *2020 IEEE Conference on Communications and Network Security (CNS)*. 2020, pp. 1–9. DOI: 10.1109/CNS48642.2020.9162207.

[13] Yi-Ting Huang et al. "Open source intelligence for malicious behavior discovery and interpretation". In: *IEEE Transactions on Dependable and Secure Computing* 19.2 (2021), pp. 776–789.

[14] Weng Qianru. *[Taiwan Information Security Conference Direct Attack] The Bureau of Investigation fully disclosed the investigation results of the ransomware attack on CPC and Formosa Plastics, and the hacker group's intrusion methods were made public*. Last accessed 20 May 2022. 2020. URL: https://www.ithome.com.tw/news/139331.

[15] William Steingartner and Darko Galinec. "Cyber threats and cyber deception in hybrid warfare". In: *Acta Polytechnica Hungarica* 18.3 (2021), pp. 25–45.

[16] *MITRE Engenuity Evaluations Trials: Deceptions APT29-Deceptions*. Last accessed 24 June 2022. 2022. URL: https://attackevals.mitre-engenuity.org/trials-deceptions/apt29-deceptions/.