Real-world detection and mitigation of AI-based cyberattacks and defence mechanisms

Mathis Durand, Yvon Kermarrec, Marc-Oliver Pahl

Real-world detection and mitigation of AI-based cyberattacks and defence mechanisms

Mathis Durand
IMT Atlantique, IRISA
mathis.durand@imt-atlantique.fr

Yvon Kermarrec
IMT Atlantique, Lab-STICC
yvon.kermarrec@imt-atlantique.fr

Marc-Oliver Pahl
IMT Atlantique, IRISA
marc-oliver.pahl@imt-atlantique.fr

***Abstract*—Cyber-physical systems are a major part of the industry. Because attacks constantly change and compromise multiple devices or components, ensuring security in these systems becomes critical. As threats to the company's systems are increasingly understood, one needs proper tools to analyze attacks or suspicious behaviors. Honeypots have existed since the eighties and evolved into different varieties of security tools, classified depending on their purpose, behavior, and architecture. Honeynets are realistic imitations of a system of information presented as an ideal target for attackers without any risk to the company. The main goal of honeynets consists of maintaining as long as possible any attacker into the fake system, capturing data such as behavior, tools, and exploits involved during the attack. When this data is collected, one can analyze it to build a more efficient defense. This paper gives a reference architecture of honeynet technology and future directions for honeynets leading to a survey. Future directions concern the legal issue of using honeypots, risks added by the implementation of honeynets, how reproducible collected attacks are, and how to motivate attackers to compromise a honeypot.**

***Index Terms*—Honeynet, Honeypot, Deceptive tool, Threat Intelligence, Data collection, Blue Team**

I. Introduction

Honeypots are systems that are used to be compromised. Using honeypots takes place in two different fields: industry and research. Because malevolent users try to break into a honeypot, a so-called blue team can collect crucial information on potential threats to a system and its potential vulnerabilities. These threats can be new exploits, new tools, or rare behavior. This can provide new solutions or highlight which parts of the system must be reconfigured. Furthermore, honeypots can distract attackers from their target. Otherwise, honeypots can be used to collect datasets and constitute valuable assets for further research and detection. As honeypots have existed since the eighties, the characterization of new generation honeynets is relevant to confront a new generation of cyber threats. Across thirty-four articles, we compared honeynet technologies of various use cases to underline commune points and discriminant.

II. Related works

As Lackner [1] explains, depending on the level of security and the information one wants to collect, honeypots can be more or less interactive. The more the honeypot is interactive, the more actions an intruder can perform. A highly interactive honeypot is a plus for data collection, but it could highly compromise a system.

Almulla et al. [2] studied data analysis in the context of SCADA systems. They also express the need for Threat intelligence data through their collection of data. Honeypot as an intrusion detection system (IDS) has been studied by Wang et al. [3]. They explain that honeypot-based IDS can reach an intrusion detection rate of 89% which is a good rate that needs to be improved. However, this system relies on Support Vector Machine training and needs a better representative dataset to improve the detection rate. This model uses a honeypot as a part of the security system and not as a tool to evaluate cybersecurity threats.

However, Mesbah et al. [4] show honeypots as a deceptive tool to reduce threats to critical infrastructure. Their paper presents an overview of the state of Operational Technology of critical infrastructure and new technologies to secure these systems such as deceptive techniques.

III. Findings

As Lackner [1] defines, a honeynet is a group of honeypots. Because attackers can execute commands from one honeypot to another, honeynets provide a highly interactive system to catch different information from classic honeypots. A honeywall is used to distinguish the actual system from multiple honeypots. Honeywalls and security systems may filter traffic and redirect users to the information system. A honeytoken is a piece of information stored in a system that looks like actual and sensitive information. It is used as bait to catch attackers and must be as realistic and valuable as possible without revealing information. Honeytokens could raise alarms when found in an unintended environment as evidence of a potential breach. Tokens must respect several principles [1] to be suitable baits and raise alarm when the honeynet has been compromised.

Honeynets are divided into three generations (see fig. 1). The first one comprises a firewall and a honeypot (1). It collects information from attackers who fail to distinguish the honeypot from the usual system.

The second generation presents a honeywall that separates malevolent traffic from actual traffic (3). A simple network (5) behind the honeywall represents an information system with highly interactive and heterogeneous honeypots. These honeypots contain honeytokens (4) to trace data, raise alarms, or build a record of interactions between honeypots. This simple network can mimic the actual information system and

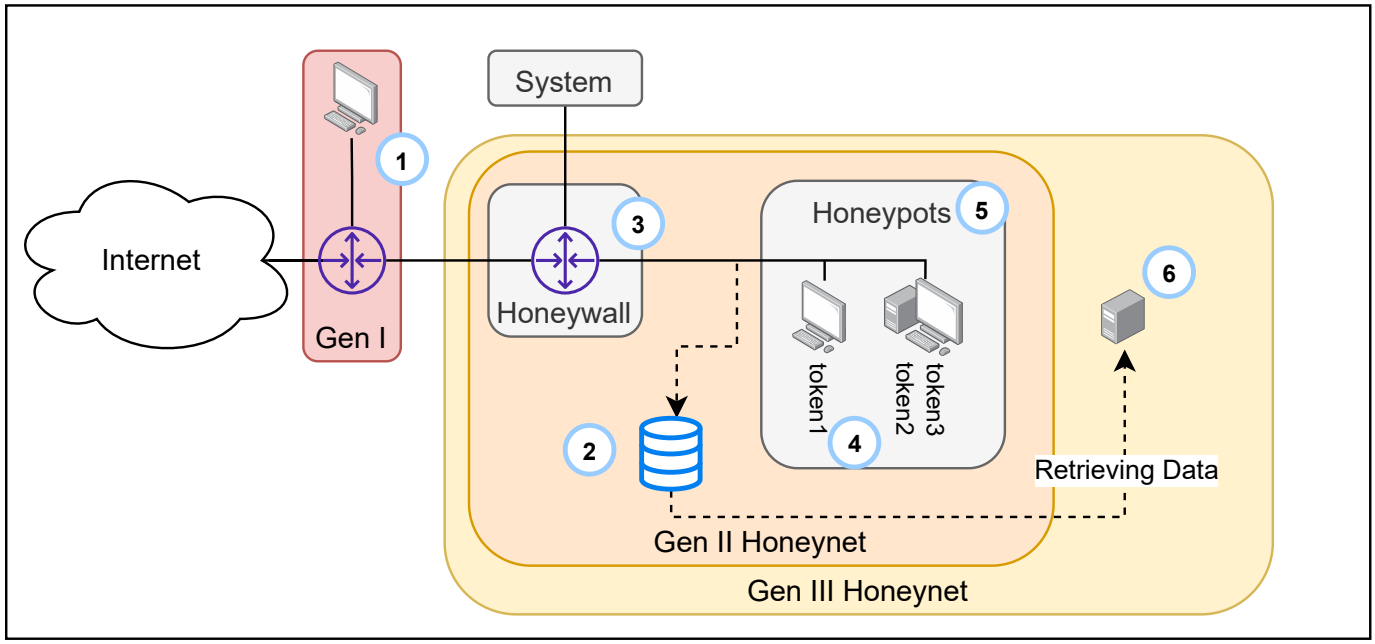

Fig. 1. Reference architecture

must be realistic to trick attackers into collecting as much information as possible. All data are collected into a dedicated server (2). Because attackers will stop their malicious activities and erase any proof of compromising when the honeypot is detected, this server must be secured and stealthy to ensure that collected data is not compromised or not representative.

The last generation adds a system (6) to monitor all data collected by a Gen II honeynet.

IV. Open issues

Honeynet must be secured enough to prevent a takeover. As Dornseif et al. [5] attests a honeypot may be detected and used to compromise the honeypot.

Collecting data raises a legal issue for honeynets as they could store attackers' data as attackers may upload data into a honeypot such as payloads or documents. The retrieved or stored data must be properly secured [6] to avoid privacy violations.

One more research question concerns reproducing attack scenarios with honeynets to test a component in a situation and collect metrics. It could measure the behavior of a security system when handling a propagating threat such as malware or ransomware.

Keeping attackers' motivation is also necessary in data collection as honeynets must be attractive to trap malevolent users. Jing et al. [7] present a model of honeynet that tries to learn the main intent of the attacker to improve the protection of the system.

One of the next steps for honeynets concerns automation and proof of trust. This combined with scalability could provide an efficient system to handle and patch new threats like zero days.

V. Conclusion

In this paper, we described the reference architecture of a third-generation honeynet, including honeypot, honeywall, and honeytoken. We also discussed the open issues of honeynets as a security tool and as a research tool. Using honeynet may increase the risks taken by the system. Catching and keeping attackers' interest is a key in honeynet effectiveness. Collecting data must ensure data privacy. Nevertheless, honeynets need automation and scalability. Our next step is the submission of a survey on current honeynets and their current and future challenges.

References

[1] P. Lackner, "How to mock a bear: Honeypot, honeynet, honeywall honeytoken: A survey," vol. 2. Science and Technology Publications, Lda, 2021, pp. 181–188.

[2] S. Almulla, C. Fachkha, and E. Bou-Harb, *Cyber Security Threats Targeting CPS Systems: A Novel Approach Using Honeypots*. [Online]. Available: https://www.researchgate.net/publication/326985688

[3] Z. Wang, J. Zhang, G. Li, Q. Liu, Y. Chi, T. Yang, and W. Zhou, "Honeynet construction based on intrusion detection." Association for Computing Machinery, 10 2019.

[4] M. Mesbah, M. S. Elsayed, A. D. Jurcut, and M. Azer, "Analysis of ics and scada systems attacks using honeypots," *Future Internet*, vol. 15, 7 2023.

[5] M. Dornseif, T. Holz, and C. N. Klein, "Nosebreak-attacking honeynets," pp. 10–11, 2004.

[6] P. Sokol, J. Míšek, and M. Husák, "Honeypots and honeynets: issues of privacy," *EURASIP Journal on Information Security*, vol. 2017, pp. 1–9, 2017.

[7] J. Tao, A. Chen, K. Liu, K. Chen, F. Li, and P. Fu, "Recommendation method of honeynet trapping component based on lstm," in *2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, 2021, pp. 952–957.