Review on AI-Driven Cyber Deception

Yash Mohite[1] Vedant Kushare[2] Yashraj Deshmukh[3] Revashree Sonawane[4] Sonali Jadhav[5]
[1,2,3,4]Student [5]Lecturer
[1,2,3,4,5]Department of Information Technology
[1,2,3,4,5]MVPS's Rajarshi Shahu Maharaj Polytechnic, Nashik, India

Abstract — In today's cybersecurity landscape, attackers continually evolve their techniques to breach systems, necessitating advanced defenses beyond traditional methods. TrapNet AI is an AI-driven cyber deception system designed to detect and mislead attackers using honeypot technology. The project targets three specific attack vectors: brute force attacks, port scanning, and SQL injection. By deploying AI models, the system analyzes traffic patterns, login attempts, and SQL queries to detect potential threats in real-time. Once an attack is detected, the system employs deception strategies, such as simulating fake vulnerabilities or introducing delays, to confuse and distract the attacker. TrapNet AI is implemented using a React and Tailwind-based frontend for user interaction, while the backend is developed with Node.js and machine learning models, handling detection and deception. The system operates within a local area network, ideal for lab environments, offering a hands-on approach to honeypot simulation. AI algorithms, including supervised and unsupervised models, are employed to distinguish between legitimate and malicious activities, providing a proactive, intelligent defense mechanism.

Keywords: Cyber Deception, Honeypot, Brute Force Attack, Port Scanning, SQL Injection, Machine Learning, AI-driven Security, Threat Detection, Network Security

I. INTRODUCTION

Nowadays, cyber-attacks advanced at a high frequency, and all these classical defense systems fail to protect against developing subtle attackers that exploit vulnerabilities due to brute-force attacks, port scanning, and SQL injections. So, there is an increased need for adaptive and proactive measures. So, TrapNet AI comes in the picture by combining artificial intelligence in cyber deception techniques using honeypot technology.

Honeypots are decoy systems that make attackers give up on their intended goals on critical assets and disclose their TTPs. TrapNet is a concept enhancement where AI-driven models get involved in the detection and response in real time regarding attacks while determining malicious activities through traffic patterns, thus creating fake vulnerabilities in real time to slowdown or mislead attackers so that the actual system remains secure.

It is frontend-based on React and Tailwind CSS while Node.js backend integrates AI models for intelligent threat detection. In this system, AI is given an integration into deception tactics for innovation in the way against cyber threats while confusing the attackers and thus enhancing overall safety of the network.

II. LITERATURE SURVEY

Chiang et al. [1] introduced the ACyDS system, the adaptive cyber deception system aimed at military real-time communications and dynamically tailoring a deception strategy so that the morale of the enemy is lowered. The system was demonstrated in 2016 at the IEEE Military Communications Conference (MILCOM), indicating the importance of adaptive deception strategies in complex threat environments. Javadpour et al. [2], reviewed and surveyed a broad number of cyber-deception techniques focusing on honeypot performance enhancement. The 2024 Computers & Security publication investigates various techniques that wit stand to harness AI integration to enhance effectiveness and realism in honeypots.

Islam and Al-Shaer [3] proposed the Active Deception Framework, a development environment to support the adaptive cyber deception ecosystem. This framework enables the security teams to extend and customize their deception strategies, presented at the 2020 IEEE Secure Development (SecDev) conference. Zhu et al. [4] carried out research in exploiting game theory and machine learning in cyber deception. Their survey, inside of the IEEE Communications Surveys & Tutorials in 2021, reports a detailed exploration of how these deep techniques stand to enhance defensive deception strategies.

Gonzalez et al. [5] presented in Algorithm Offense and Defense in the Cyber-Physical Environment: An Integrated Systems Approach, Adaptive cyberdefense with deception: Integrating artificial intelligence and human reasoning, where an Advanced Cyber Denial and Deception Cognitive Framework is proposed briefly in their 2022 book chapter on cyber deception. The work addresses how artificial intelligence is being used to serve humans in making deception decisions much more effectively. Alternatively, Kouremetis et al. [6] has developed Mirage in-2015, A framework for cyber deception platforms as an autonomous defense against sophisticated attackers. The details on the use of Mirage's emulation, for deceiving the attackers and allowing for instant detection and reaction to a wide range of extreme threats have been presented, by the authors in their 2024 Annals of Telecommunications study.

On the other another hand, Bharadiya [7] worked on the role of machine learning in cyber security. The paper reviews the latest research in this field and discusses how machine learning model are integrated with the cyber security systems, which is vital in gaining an edge against the growing threats. At the same time, another work by Achleitner et al. [8] was also presented at MIST '16, Cyber Deception using virtual networks to suppress insider reconnaissance, A low-interaction honeypot trapping Robocop-attack agent composed in 2013 carried out by a research group, using virtualized deception environments to overcome insider threats.

In their paper published in Computers & Security, Zhang and Thing [9] provide a comprehensive review of three decades of deception techniques used in active cyber defense. The authors retrospectively analyze the evolution of deception-based methods, focusing on their application in

enhancing cybersecurity defenses. They explore how these techniques have been employed to mislead attackers and gather valuable intelligence for defending networks. Furthermore, the paper offers an outlook on future developments in this field, emphasizing the growing importance of integrating deception with automated and AI-driven approaches for more effective cyber defense strategies.

Cai and Koutsoukos [10] continue the discussion of real-time detection of deception attacks in cyber-physical systems with a focus on how these attacks can be detected and mitigated in critical infrastructure. Their paper, published in 2023 in the International Journal of Information Security, investigates the integration of real-time data analysis into deception for effective defense.

Lastly, Urias et al. [11] at the 2016 IEEE Symposium on Technologies for Homeland Security. The authors dealt with the use of deception to gain intelligence from adversaries, which is supposed to improve their overall network defense. This idea is based on the understanding of the behavior of attackers in controlled environments.

III. PROPOSED SYSTEM

TrapNet AI is one of the artificial intelligence-based cyber deception systems that are used for the purpose of both the detection and diversion of the attackers who attack local area networks (LAN). It uses machine learning models to detect real-time attack vectors like brute force attacks, port scanning, and SQL injection. The abstract model of the system is a React and Tailwind CSS-based frontend that is used for monitoring and configuration. On the other side, the backend is created with Node.js, which does the management of traffic analysis by the help of a combination of supervised and unsupervised models. This scheme integrates legitimate and malicious activities, thereby offering an ability to distinguish between them, and it can also pinpoint the particular threat. When TrapNet AI detects a likely attack, it works by activating fake honeypots that are actually the imitation of vulnerable systems diverting the aim of the attackers, and at the same time, they collect observations about the attackers' behaviors.

The system is not limited to simple detection only, as it employs the use of advanced deception strategies, like the invention of artificial delays, and provision of false data, by it doing so, the big advantage of this is that it confuses the attackers making them think that they have reached their goal when in reality the real assets are still untouched. Divider AI is a product of the division of language that is completely controlled inside a LAN environment and is used more for purposes of education in cybersecurity. The system records every interaction through the logging facility, thus, generating comprehensive reports on the post-attack analysis. The system that is developed through scalability and future integration with the threat intelligence platforms to be equipped with features that would enable the detection of the threats beforehand and the use of the adaptive deception to outmaneuver the attackers, which would lead to network security being enhanced, is the TrapNet AI.

IV. SYSTEM ARCHITECTURE

The system architecture of TrapNet AI is built to provide a robust cybersecurity solution through AI-driven cyber deception. The core of the architecture is the Admin Application, responsible for monitoring traffic, detecting threats with AI models, deploying adaptive honeypots, and maintaining logs of attacks. The Traffic Monitoring module continuously scans network activity, while the AI Detection Models examine patterns to detect any malicious attacks. If a threat is detected, Honeypot Systems will create decoys that lure attackers away from critical assets and gather valuable intelligence in the process. The Logs and Report module records all the detected threats for further analysis.

The User Application offers a User Workspace where users can manage settings and receive alerts about potential threats. This is supported by the database, where Attack Logs, Attack Data, and User Information are maintained in a centralized record for monitoring and analysis. The Visualization Module allows administrators to see system performance in visual form and track key metrics for optimizing defenses. All these parts ensure that TrapNet AI will provide an adaptive, scalable solution toward real-time detection, analysis, and response against cyber threats.

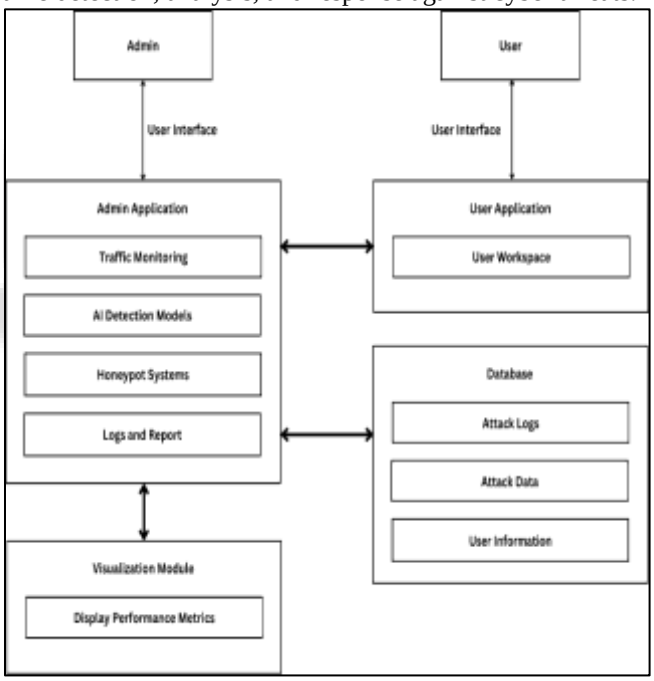

Fig. 1: Architecture

V. FEATURES OF THE SYSTEM

A. AI-Driven Threat Detection:

The system has AI-enabled detection of malicious activities actively using machine learning algorithms to detect different kinds of malicious activities, including brute-force attacks, port scanning, and SQL injection. TrapNet AI finds a particular threat at a high level of accuracy by analyzing anomalies and traffic patterns in real-time.

B. Dynamic Honeypot Deployment:

TrapNet AI simulates vulnerable services or systems using dynamic honeypots that are deployed whenever an attack is detected. These honeypots fake systems with real asset

characteristics, tantalizing attackers with false vulnerabilities and thereby diverting their efforts away from operational systems.

C. Supervised and Unsupervised Learning Models:

The system offers both supervised learning (to detect known attack patterns) and unsupervised learning (to detect new, unusual behaviors). The two-pronged technique guarantees the detection of broad-spectrum as well as novel threats.

D. Real-Time Response and Deception:

TrapNet AI introduces artificial delays in channels of communication with the detection of detected threats in real-time, feeding misinformation to its attackers. This is important in trying or buying time to confuse its attackers while keeping the system intact.

E. User-Friendly Interface:

The system has a front-end based on React and Tailwind CSS that provides an easy-to-use dashboard for administrators. The interface features real-time traffic data monitoring, system logs, and customizable honeypot configurations.

F. Attack Behavior Logging:

Interactions with the honeypots by attackers are extensively logged. It captures such points as attack payloads, IP addresses, and sequences of commands. This data is useful for post-attack analysis for further system upgrading and enhancements.

G. Scalability and Flexibility:

While TrapNet AI was originally designed to penetrate lab environments, it can easily be scaled in its future versions to stand for scalability in supporting large, enterprise network environments. No less important, it can integrate with external threat intelligence platforms and keep itself up-to-date on possible attack vectors, human, or automated.

VI. Conclusion

Therefore, TrapNet AI conclusively demonstrates the possibility of using AI-driven cyber deception to reveal and counter attacks such as brute-force attempts, port scanning, and SQL injection. This is indeed where the actual potential of machine learning models meets a honeypot system with the defense mechanism that can adapt in real-time to identify an attack while simultaneously engaging attackers through deception, minimizing damage, and gathering intelligence about their techniques.

This research is crucial in augmenting value for the emerging research field of AI-based cybersecurity by offering a practical solution to modern cyber defense strategies that can outsmart traditional detection methods through deception-based scenarios that impede attackers and enhance security within networks.

VII. Acknowledgment

We are deeply thankful to our project guide Ms. S. S. Jadhav for their continuous support, guidance, and encouragement throughout the project. Their valuable insights and feedback helped us refine our ideas and bring this project into actualization

References

[1] C.-Y. J. Chiang et al., "ACyDS: An adaptive cyber deception system," MILCOM 2016 - 2016 IEEE Military Communications Conference, Baltimore, MD, USA, 2016, pp. 800-805, doi: 10.1109/MILCOM.2016.7795427.

[2] A. Javadpour, F. Ja'fari, T. Taleb, M. Shojafar, and C. Benzaïd, "A comprehensive survey on cyber deception techniques to improve honeypot performance," Computers & Security, 2024, doi: 10.1016/j.cose.2024.103792.

[3] M. M. Islam and E. Al-Shaer, "Active Deception Framework: An Extensible Development Environment for Adaptive Cyber Deception," 2020 IEEE Secure Development (SecDev), Atlanta, GA, USA, 2020, pp. 41-48, doi: 10.1109/SecDev45635.2020.00023.

[4] M. Zhu, A. H. Anwar, Z. Wan, J.-H. Cho, C. A. Kamhoua, and M. P. Singh, "A Survey of Defensive Deception: Approaches Using Game Theory and Machine Learning," IEEE Communications Surveys & Tutorials, vol. 23, no. 4, pp. 2460-2493, Fourthquarter 2021, doi: 10.1109/COMST.2021.3102874.

[5] Gonzalez, P. Aggarwal, E. A. Cranford, and C. Lebiere, "Adaptive Cyberdefense with Deception: A Human–AI Cognitive Approach," in Cyber Deception, Springer, 2022, pp. 41-57, doi: 10.1007/978-3-031-16613-6_3.

[6] M. Kouremetis, D. Lawrence, R. Alford et al., "Mirage: cyber deception against autonomous cyber-attacks in emulation and simulation," Annals of Telecommunications, 2024, doi: 10.1007/s12243-024-01018-4.

[7] J. Bharadiya, "Machine Learning in Cybersecurity: Techniques and Challenges," European Journal of Technology, vol. 7, pp. 1-10, 2023, doi: 10.47672/ejt.1486.

[8] S. Achleitner, T. La Porta, P. McDaniel, S. Sugrim, S. V. Krishnamurthy, and R. Chadha, "Cyber Deception: Virtual Networks to Defend Insider Reconnaissance," in Proceedings of the 8th ACM CCS International Workshop on Managing Insider Security Threats (MIST '16), New York, NY, USA, 2016, pp. 57–68, doi: 10.1145/2995959.2995962.

[9] L. Zhang and V. L. L. Thing, "Three decades of deception techniques in active cyber defense - Retrospect and outlook," Computers & Security, vol. 106, 2021, doi: 10.1016/j.cose.2021.102288.

[10] F. Cai and X. Koutsoukos, "Real-time detection of deception attacks in cyber-physical systems," International Journal of Information Security, vol. 22, pp. 1099-1114, 2023, doi: 10.1007/s10207-023-00677-z.

[11] V. E. Urias, W. M. S. Stout, and H. W. Lin, "Gathering threat intelligence through computer network deception," 2016 IEEE Symposium on Technologies for Homeland Security (HST), Waltham, MA, USA, 2016, pp. 1-6, doi: 10.1109/THS.2016.7568916.

[12] M. AbuOdeh, C. Adkins, O. Setayeshfar, P. Doshi, and K. H. Lee, “A Novel AI-based Methodology for Identifying Cyber Attacks in Honey Pots”, AAAI, vol. 35, no. 17, pp. 15224-15231, May 2021

[13] Sarker, Iqbal & Furhad, Md & Nowrozy, Raza. (2021). "AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions." SN Computer Science. 2. doi: 10.1007/s42979-021-00557-0.
[14] P. Lanka, K. Gupta, and C. Varol, "Intelligent Threat Detection—AI-Driven Analysis of Honeypot Data to Counter Cyber Threats," Electronics, vol. 13, no. 13, p. 2465, 2024, doi: 10.3390/electronics13132465.
[15] B. Olafuyi, "Artificial Intelligence in Cybersecurity: Enhancing Threat Detection and Mitigation," International Journal of Scientific and Research Publications, vol. 13, pp. 194-200, 2023, doi: 10.29322/IJSRP.13.12.2023.p14419.
[16] Nand, Kumar, Et, al.. (2023). "AI in Cybersecurity: Threat Detection and Response with Machine Learning.", vol 44 No.33,doi: 10.52783/tjjpt.v44.i3.237.
[17] Daniel Zielinski, Hisham A. Kholidy (2022) "An Analysis of Honeypots and their Impact as a Cyber Deception Tactic" doi:10.48550/arXiv.2301.00045
[18] Dr. Nirvikar Katiyar, Mr. Somendra Tripathi, Mr. Praveen Kumar, Mr. Shekhar Verma, Dr. Alok Kumar Sahu, & Dr. Shailesh Saxena. (2024). "AI and Cyber-Security: Enhancing threat detection and response with machine learning". Educational Administration: Theory and Practice, 30(4), 6273–6282. doi:10.53555/kuey.v30i4.2377
[19] M. Lourens, A. P. Dabral, D. Gangodkar, N. Rathour, C. N. Tida and A. Chadha, "Integration of AI with the Cybersecurity: A detailed Systematic review with the practical issues and challenges," 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 1290-1295, doi: 10.1109/IC3I56241.2022.10073040.
[20] Basiru A. Olafuyi (2023); Artificial Intelligence in Cybersecurity: Enhancing Threat Detection and Mitigation; International Journal of Scientific and Research Publications (IJSRP) 13(12) (ISSN: 2250-3153), doi:10.29322/IJSRP.13.12.2023.p14419

