IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI-Driven Honeypot System For Proactive Detection And Defense Against Cryptographic Ransomware

[1]Parthiban B, [2]Lavanya J, [3]Naumaan Faaize M, [4]Mohamed Sameer M, [5]Abdulsha Ashiq F
[1,3,4,5] Student, [2]Assistant Professor
[1]Department of Computer Science and Engineering,
[1]Aalim Muhammed Salegh College of Engineering, Chennai, India

Abstract: The proliferation of cryptographic ransomware poses a severe threat to organizational data security, encrypting critical assets and demanding payments with no assurance of recovery. Conventional defenses, reliant on signature- based detection, falter against rapidly evolving ransomware variants, necessitating innovative solutions. This paper presents an AI- driven honeypot system integrated with a Bidirectional Long Short-Term Memory (BiLSTM) model and Format Preserving Encryption (FPE) to proactively detect and defend against ransomware. The system employs dynamic honeypots to lure ransomware, capturing attack patterns to enhance the BiLSTM model's real- time detection capabilities. FPE ensures data protection by camouflaging files, maintaining accessibility for legitimate users while thwarting unauthorized encryption.

Index Terms **– Cryptographic Ransomware, AI-Driven Honeypot, Bidirectional Long-Short term Memory (BiLSTM).**

I. Introduction

Ransomware, a malicious software that encrypts or locks victims' data to extort payments, has emerged as a critical cybersecurity challenge. Leveraging cryptocurrencies for anonymous transactions, ransomware targets sensitive sec tors, causing financial losses and operational disruptions. Traditional antivirus solutions, such as those from Norton and Kaspersky, pri marily utilize signature-based detection, which identifies malware based on known features. However, this approach struggles to counter new or polymorphic ransomware variants, leav ing organizations vulnerable. The need for adap tive, real-time defense systems has become paramount.

This paper proposes an AI-driven honeypot sys tem that integrates advanced machine learn ing and encryption techniques to address these challenges. By combining dynamic honeypots, BiLSTM-based detection, and FPE, the system offers a proactive approach to ransomware defense, building on the concept of HoneyFiles to create a robust, evolving security framework. The architecture of this system is further reinforced by the concept of HoneyFiles—strategically placed decoy files designed to trigger alerts upon unauthorized access—creating a multilayered and proactive defense mechanism. To address these limitations, this paper introduces an artificial intelligence (AI)-driven honeypot system designed to enhance ransomware detection and prevention. The proposed system integrates dynamic honeypots—deceptive computing environments that lure and monitor attacker behavior—with advanced deep learning models based on Bidirectional Long Short-Term Memory (BiLSTM) networks for accurate behavioral analysis. Furthermore, it incorporates Format-Preserving Encryption (FPE) to secure sensitive data while maintaining operational usability. The integration of these technologies builds upon the concept of HoneyFiles—decoy files designed to detect unauthorized access—resulting in a robust and adaptive security infrastructure.

This proactive and intelligent approach aims to outpace the evolving threat landscape and provide a comprehensive defense against ransomware attacks. Through this integrated approach, the proposed framework not only detects ransomware activity with high precision but also adapts over time by learning from emerging threat patterns. This research contributes a scalable, intelligent, and resilient solution to the field of cybersecurity, offering a forward-looking strategy for ransomware detection, analysis, and prevention. To mitigate this growing threat, the need for intelligent, real-time, and adaptive defense mechanisms has become paramount. In response to these challenges, this paper presents a novel AI-driven honeypot system that leverages a combination of dynamic deception technologies, deep learning-based behavioral analysis, and data obfuscation techniques. While effective for previously identified threats, these tools exhibit a critical weakness: they fail to detect new, unknown (zero-day), or polymorphic ransomware strains that can mutate their code to bypass standard defenses. As attackers continually refine their tactics, this static defense model becomes increasingly inadequate.

II. Background

The use of ransomware as a cyber weapon has surged in recent years, largely due to the profitability and anonymity it affords attackers. Ransomware typically gains access to a system through phishing emails, malicious downloads, or unpatched software vulnerabilities. Once inside, it rapidly encrypts files and renders them inaccessible, displaying a ransom note that demands payment—usually in cryptocurrency—to unlock the data. Despite various efforts to combat this threat, traditional defense methods have repeatedly shown limitations. Signature-based detection techniques depend on previously identified malware samples to flag threats. While effective against known malware, these systems are unable to detect zero-day attacks or polymorphic ransomware that frequently changes its code to evade detection.

Moreover, existing malware analysis tools are often not integrated and rely heavily on human analysts, making them slow and inefficient. Automation is minimal, and there is a general lack of intelligence in how these systems respond to new threats. To address these deficiencies, the cybersecurity field has seen increased interest in deception technologies like honeypots. Honeypots are decoy systems designed to mimic real operational environments, thereby attracting attackers and capturing their methods without putting actual assets at risk. When augmented with artificial intelligence and machine learning, honeypots evolve into dynamic systems capable of learning from interactions with malicious software.

This paper leverages that concept, enhancing it with BiLSTM models that offer deep temporal analysis of ransomware behavior. Furthermore, by integrating Format Preserving Encryption, the system provides a last line of defense that ensures critical data remains unreadable to ransomware, thereby mitigating the consequences of a successful attack. Together, these components form a holistic approach that not only identifies and contains ransomware but also adapts to new threats as they emerge. The sophistication of the BiLSTM model contributes significantly to this framework. Unlike conventional models that process input in a single temporal direction, BiLSTM networks analyze input sequences both forward and backward. This dual temporal comprehension allows the model to identify complex patterns and interdependencies in system behavior that may indicate malicious intent.

Ransomware, which often follows identifiable sequences—such as gaining elevated privileges, accessing and modifying files, and establishing external communications—can be effectively detected by such temporal pattern analysis. As the honeypot records these activities, the BiLSTM learns from both successful and failed attacks, continually refining its ability to discern between normal user activity and potential ransomware behavior. This creates a feedback loop where the detection model becomes increasingly accurate over time, reducing false positives and improving the speed and reliability of alerts.

Traditional systems, which focus on known indicators of compromise (IOCs), are inherently reactive and fail to offer the level of foresight needed to combat ransomware in real-time. In contrast, AI-enhanced honeypots act as active participants in the security framework. They do not passively wait for attacks to happen; instead, they invite interaction by replicating real network assets and services, thus tricking malicious entities into revealing their strategies and tools. . Even in scenarios where ransomware successfully infiltrates a network and bypasses initial detection layers, FPE ensures that the files targeted by the malware are encrypted in a manner that maintains their format but renders them useless to unauthorized users. This frustrates the ransomware's objective by denying it access to data it can effectively encrypt or manipulate.

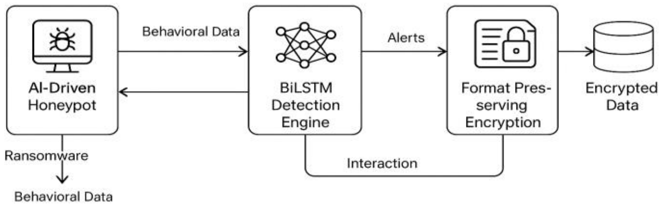
III. Architecture

The architecture of the proposed AI-driven honeypot system is designed to provide layered and intelligent defense against cryptographic ransomware. At the core of this system are three tightly integrated components: dynamic honeypots, a BiLSTM-based detection engine, and a Format Preserving Encryption module. The dynamic honeypots are the first point of interaction for ransomware. These are highly configurable decoy systems that simulate real-world vulnerabilities and sensitive file structures to entice ransomware. Once engaged, the honeypots record comprehensive behavioral data including file access logs, process creation patterns, and network activity.

This raw data is then transmitted in real-time to the BiLSTM detection engine. The BiLSTM model processes the data bidirectionally, analyzing it in both forward and backward time steps to uncover complex patterns indicative of ransomware activity. This capability is particularly valuable for identifying new or obfuscated variants of ransomware, as the model does not rely on static signatures but instead learns from behavioral sequences. Upon detection of ransomware-like activity, the system can initiate containment protocols and alert administrators.

Parallel to this detection mechanism, the Format Preserving Encryption module ensures that critical user data remains encrypted in a format that mirrors the original, making it difficult for ransomware to identify and encrypt these files. This form of encryption is strategically important because it allows legitimate users to access data without delay or compatibility issues, while simultaneously thwarting ransomware efforts to compromise the information. The architecture is further supported by a central monitoring and control dashboard that aggregates logs, issues alerts, and provides configuration options for dynamic honeypot deployment and BiLSTM retraining. This architectural synergy enables a proactive, intelligent, and resilient defense system that evolves continuously in response to new cyber threats.

Figure 1: System Architecture



Key Components of the Proposed System

A. *AI-Driven Honeypot Module:*

The AI-driven honeypot module constructs highly sophisticated decoy environments that emulate real enterprise systems, such as file servers, databases, or endpoints, with remarkable accuracy to lure ransomware. These decoys replicate vulnerable assets by mimicking realistic file structures, network traffic patterns, and system behaviors, using machine learning to dynamically adapt their characteristics based on observed attacker tactics, ensuring they remain attractive to evolving threats. When ransomware interacts with the honeypot, the module captures detailed behavioral data in real-time, including file access logs that record timestamps, file types, and access frequency to reveal targeting strategies, process behavior that monitors spawned processes, resource consumption, and inter-process communication to identify malicious patterns, and system call sequences that expose low-level operations like file operations or encryption routines. This data is collected within a sandboxed environment to isolate the ransomware, preventing escape, and is fed into a centralized pipeline for analysis. Reinforcement learning optimizes decoy configurations over time, enhancing effectiveness by analyzing past interactions, making the captured data a critical input for the detection engine to proactively identify and anticipate new ransomware variants.

B. *BiLSTM Detection Engine:*

The Bidirectional Long Short-Term Memory (BiLSTM) detection engine employs a deep learning architecture to analyze system behavior by processing temporal sequences in both forward and backward directions, offering a comprehensive understanding of temporal context unlike traditional unidirectional models. It processes inputs from the honeypot, including file access logs, process behavior, and system call sequences, to detect ransomware through temporal patterns like rapid file modifications followed by encryption calls, contextual relationships such as processes accessing sensitive files while communicating with external servers, and behavioral anomalies like unusual process spawning or system call frequencies. Trained on diverse datasets of ransomware and benign behaviors, the BiLSTM generalizes across known and unknown threats, using bidirectional analysis to capture subtle indicators like preparatory reconnaissance that unidirectional models might overlook.

C. *Format Preserving Encryption (FPE):*

Format Preserving Encryption (FPE) cryptographically secures sensitive data while preserving its original format, structure, and usability, ensuring that encrypted data retains the same length, character set, and data type as the plaintext, such as a 16-digit credit card number encrypting into another 16-digit number or an email address remaining a valid string. This technique thwarts ransomware by obfuscating data, making it unrecognizable and unencryptable by attackers, while allowing legitimate applications and users to interact with encrypted data without decryption, maintaining compatibility with existing workflows like database queries or file processing. Implemented using NIST-approved FF1 or FF3-1 algorithms, FPE supports customizable encryption domains and is integrated at the file-system or database level to automatically encrypt sensitive data upon creation or modification.

D. *Enterprise System Interface and Response Layer:*

The Enterprise System Interface and Response Layer integrates the ransomware defense system with enterprise infrastructure, including servers, endpoints, databases, and cloud environments, facilitating seamless communication between AI-driven components and organizational assets for real-time threat management. It generates prioritized alerts with detailed metadata like affected systems, ransomware type, and confidence scores for rapid decision-making, provides interactive visualization dashboards with heatmaps, timelines, and graph-based representations of system health, honeypot activity, and detection metrics, and executes automated responses like isolating infected endpoints, terminating malicious processes, or restoring files from backups based on detection confidence.

E. *Workflow and Integration:*

The system’s workflow orchestrates a seamless progression from honeypot engagement to detection and encryption, designed to trap, analyze, and neutralize ransomware automatically. It begins with the honeypot luring attackers, capturing behavioral data, which the BiLSTM engine analyzes to detect threats, followed by FPE securing critical data to block ransomware encryption. Each step integrates smoothly to ensure minimal disruption and maximum protection, leveraging real-time data flows and automated responses to maintain enterprise continuity while neutralizing threats, with the system adapting to evolving ransomware tactics through continuous learning and optimization.

IV. Conclusion

The AI-driven honeypot system presented in this paper offers a comprehensive solution to the growing threat of cryptographic ransomware. By integrating dynamic honeypots, BiLSTM-based detection, and FPE, the system achieves real- time threat identification and data protection, surpassing the limitations of signature-based defenses. Its autonomous operation and adapt- ability ensure resilience against evolving ransomware variants, while its seamless integration into existing systems enhances scalability. Future work will focus on improving detection precision in complex network settings and exploring additional machine learning architectures to further enhance performance.

This research contributes to cybersecurity advancements by providing a robust, innovative framework for ransomware defense, safeguard- ing critical data and organizational operations. Format-Preserving Encryption adds another layer of resilience by encrypting data in a way that maintains its original structure and format. This approach ensures that sensitive information is protected without disrupting the functionality of the system or applications that rely on specific data formats. It allows the system to uphold usability while maintaining strong data security. Designed for autonomous operation, the system requires minimal human oversight and is capable of continuously adapting to new threats. Its compatibility with existing IT infrastructures makes it a scalable solution that can be deployed across diverse organizational environments.

Looking ahead, future research will focus on refining detection accuracy in large-scale and complex network ecosystems, where the volume and variety of data can challenge existing models. Moreover, the exploration of alternative machine learning architectures may yield further enhancements in detection efficiency and overall system performance.

References

[1] H. Oz, A. Aris, A. Levi, and A. S. Uluagac, “A survey on ransomware: Evolution, taxonomy, and defense solutions,” ACM Computing Surveys, vol. 54, no. 11s, pp. 1–37, Jan. 2022, doi: 10.1145/3514229.

[2] D. Smith, S. Khorsandroo, and K. Roy, “Machine learning algorithms and frameworks in ransomware detection,” IEEE Access, vol. 10, pp. 117597–117610, Nov. 2022, doi: 10.1109/ACCESS.2022.3218779.

[3] T. McIntosh, M. Blowers, and C. Arndt, “Ransomware mitigation in the modern era: A systematic literature review,” ACM Computing Surveys, vol. 55, no. 9, pp. 1–36, Dec. 2022, doi: 10.1145/3543876.

[4] A. Kapoor et al., “Ransomware detection, avoidance, and mitigation scheme: A review and future directions,” Sustainability, vol. 14, no. 1, p. 8, Dec. 2021, doi: 10.3390/su14010008.

[5] Q. Chen and R. A. Bridges, “Automated behavioral analysis of malware: A case study of WannaCry ransomware,” in Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA), pp. 454–460, Dec. 2022, doi: 10.1109/ICMLA.2022.0-119.

[6] G. Cusack, O. Michel, and E. Keller, “Machine learningbased detection of ransomware using SDN,” in Proc. ACM Int. Workshop on Security in Software Defined Networks & Network Function Virtualization, 2022, pp. 1–6, doi: 10.1145/3180465.3180467