AI-Powered Ransomware Detection: A Comprehensive Survey on Machine Learning and Deep Learning Techniques

Muhammad Junaid Iqbal and Jordi Serra-Ruiz

Universitat Oberta de Catalunya (UOC),
CYBERCAT-Center for Cybersecurity Research of Catalonia,
Barcelona, Spain

Abstract. Ransomware has emerged as a critical and rapidly evolving cybersecurity threat, significantly impacting sectors such as healthcare, finance, and government infrastructures. This paper presents a comprehensive survey of contemporary ransomware detection techniques, focusing on machine learning (ML) and deep learning (DL) methodologies, which have shown promise in adapting to the rapidly changing landscape of ransomware attacks. The survey includes a detailed comparative analysis of static, dynamic, and hybrid detection models, highlighting their respective advantages and limitations. The key findings from the survey show that ML and DL-based methods have a better detection capabilities but still having challenges such as large and diverse datasets, the computational cost of advanced techniques, and model adaptability across various platforms still exist. We also delve into some up-and-coming trends, like quantum computing and federated learning, both of which have the potential to overcome present limitations in computation efficiency and privacy concerns, respectively. It also points to the increasing attention being paid to adversarial defenses, which seek to make models more robust against complex evasion attempts.

Keywords: Ransomware; Machine learning; Deep Learning; Cybersecurity; Hybrid Detection; Adversarial Defense

1 Introduction

Traditional methods of detection against ransomware, such as signature-based techniques and heuristic analysis, are becoming ineffective in the fight against these sophisticated variants of ransomware. Signature-based detection relies on predefined patterns and signatures; thus, it fails to detect new and unknown ransomware malware, which is ever-evolving. The malware developers usually use techniques such as code obfuscation and polymorphism, where the malware dynamically changes its structure to avoid detection. Further, sandbox evasion is a technique through which ransomware identifies whether it runs in a controlled environment and delays its payload activation to avoid detection during the analysis. Heuristic analysis, on the other hand, is more flexible but depends on identifying behavioral anomalies and often struggles to differentiate between legitimate and malicious activities when ransomware may masquerade as normal system processes. These limitations indicate that advanced detection methodologies are needed to identify the dynamic and adaptive nature of ransomware in modern attacks. Several works have highlighted the potential for enhanced detection by incorporating machine learning and deep learning to increase accuracy, considering patterns that traditional methods might not find [1].

Cybersecurity experts have begun to explore machine learning (ML) and deep learning (DL) as viable alternatives to improve ransomware detection. These ML and DL techniques allow systems to analyze large datasets and recognize patterns that traditional methods might miss. A significant benefit of ML and DL is their capability to identify zero-day ransomware variants that have not been previously detected by existing signature-based

David C. Wyld et al. (Eds): COMIT, AISO, CRBL, WIMNET, EDUPT– 2025
pp. 47-63, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.151204

systems. These methods are more flexible in responding to changing ransomware threats, as they can continuously learn and adapt to new attack patterns without requiring manual updates.

Various machine learning approaches, including supervised learning algorithms, like decision trees, random forests, and support vector machines have been applied successfully in ransomware detection. These techniques will analyze features extracted from suspicious files, system behavior, and network traffic to classify whether a given sample is malicious [1]. Deep learning, a subcategory of ML, has proved even more promising by utilizing complex architectures such as Convolutional Neuronal Network (CNN) and Recurrent neural networks (RNN). These models can automatically identify high-level features from the raw data and therefore have turned out to be particularly good at finding patterns within large unstructured datasets [2]. While ML and DL offer exciting potential, several challenges remain. Among these, the most significant bottleneck is the general lack of comprehensive, diverse, and up-to-date ransomware datasets that are mandatory for training robust models. In addition, most of the publicly available datasets on ransomware are either obsolete, limited in size, or relevant to only certain varieties of ransomware, which eventually creates obstacles in the development of generalized detection models [3]. Many ML/DL-based techniques are computationally demanding, requiring numerous resources for processing and analyzing real-time data. This limitation is particularly critical in resource-constrained environments, such as mobile devices or edge computing systems, where processing power and memory are limited [4]. Adversarial attacks, where ransomware is deliberately designed to evade detection models, present a growing concern, requiring advanced techniques such as adversarial training to improve model resilience [2].

1.1 Contributions and organization of the paper

The rest of the paper is organized as follows: Section 2 presents a discussion of ransomware taxonomy and traditional detection methods. Section 3 outlines recent methods of machine learning and deep learning to detect ransomware, including main algorithms and frameworks. Section 4 discusses the performances of these methods, further explaining their advantages and disadvantages, and suitability for different platforms. Section 5 provides a discussion of the comparative analysis of ransomware detection methods Finally, Section 6 discusses the challenges faced by ransomware detection and future research directions including adversarial defense, quantum computing, and cross-platform detection frameworks.

2 Background and Related Work

This section provides the necessary background ideas that serve as the underpinning for ransomware detection to be comprehensible. It starts by presenting a ransomware taxonomy before presenting an overview of classical detection methods and accompanying research in this area.

2.1 Taxonomy of Ransomware

Ransomware has evolved significantly over the years, and as such, it can be broadly categorized into three major types based on their operational characteristics: locker ransomware, crypto-ransomware, and hybrid ransomware. Each type employs unique methods to achieve its goal-typically extorting money from the victim.

- **Locker Ransomware:** This type locks users out of their systems, often displaying a ransom demand for restored access. Examples include `Reveton and Winlocker`, which block system functionality until a payment is made [5].
- **Crypto-Ransomware:** This more prevalent and harmful variant encrypts user files, rendering them inaccessible without a decryption key. Prominent examples include WannaCry, `Ryuk, and Conti,` which use robust cryptographic methods, making recovery nearly impossible without paying the ransom [3].
- **Hybrid Ransomware:** Combining features of both locker and crypto-ransomware, hybrid ransomware locks systems and encrypts files simultaneously. Notable instances like `Petya and GoldenEye` have caused significant disruptions in businesses and government entities[6].

The evolution of ransomware types has made it increasingly difficult for traditional security measures to identify and mitigate these threats effectively. Therefore, advanced detection methods, especially those based on Machine learning and deep learning, have become crucial in the identification and response to these evolving threats.

2.2 Traditional Detection Techniques

Traditional methods for ransomware detection involved static and dynamic analyses. In both techniques, certain merits and severe limitations were noticed regarding sophisticated and modern ransomware strains.

Static Analysis: Static analysis examines the code or binary of a suspicious file without executing it. Known ransomware can often be detected using this method by matching the file's signature with those stored in a database. Although static analysis is fast and non-invasive, it can be easily circumvented through techniques such as polymorphism or code obfuscation, wherein the ransomware's code changes with each execution to evade detection methods based on signature matching [2].

Dynamic Analysis: Unlike static analysis, dynamic analysis involves running the suspected ransomware in a controlled, isolated environment—such as a sandbox—while actively monitoring its behavior. This approach is capable of identifying newly developed ransomware variants that exhibit typical malicious activities, including file encryption, registry modifications, and unauthorized network communications. Dynamic analysis demands significant computational resources and is susceptible to evasion techniques employed by advanced ransomware strains. For instance, some ransomware is specifically designed to detect when it is running in a virtual machine or sandbox environment, delaying the activation of its payload until it confirms the presence of a real system [4].

While both static and dynamic analysis-based approaches are highly limited against the polymorphic, metamorphic, and fileless ransomware attack variants, this necessitates further adaptive intelligent detection techniques, mostly based on ML and DL.

Intrusion Detection and Prevention Systems (IDS/IPS). Conventional security network systems like Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS) are typically used to identify and respond to many types of cyberattacks. These systems make use of network traffic monitoring and comparison of patterns with a database of attack signatures or behavior-based rules. IDS/IPS has worked well in detecting traditional attacks but can only make limited responses to contemporary ransomware attacks. Ransomware frequently employs encrypted or obfuscated payloads, polymorphic behavior, and new delivery vectors that don't trigger traditional IDS/IPS sensing. Intrusion Detection and Prevention Systems solutions are largely network-oriented, whereas ransomware

engages endpoint activity and system changes extensively, which need more in-depth inspection offered by machine learning and deep learning-based approaches. Therefore, although IDS/IPS offer baseline security, they are frequently inadequate alone to detect and counter advanced ransomware assaults and require complementary solutions and controls.

2.3 Related Work

Various studies have explored the potential of machine learning and deep learning to enhance ransomware detection. These approaches provide systems that can learn from data and identify patterns in both known and unknown variants. These methods greatly improve over the traditional approaches.

Opcode Analysis: The opcode analysis framework for the early detection of ransomware has been proposed by Pesem et al., [3]. This approach provides a more profound understanding of ransomware behavior through examining low-level operational instructions in system memory. Hence, it enables both early detection and attribution of ransomware families. This method achieves scalability and can be applied in real-time detection scenarios, making it well-suited for a wide variety of operational environments. It still requires extensive datasets of opcode sequences for training, which might be challenging to acquire.

RANSOMNET+: Singh et al. [2] proposed the `RANSOMNET+` system, a deep learning model that leveraged CNNs and transformers for the extraction of hierarchical and local patterns from cloud-encrypted data. This system has demonstrated superior performance in cloud environments with precisions and recalls over 98%. Although very efficient, the major setback for RANSOMNET+ is that it is very computationally expensive, thus requiring high hardware resources both for training and inference. Consequently, its application in mobile devices or IoT systems might be restricted.

DeepWare: Ganfure et al.[4] developed `DeepWare`, a system that leverages hardware performance counters (HPC) to detect ransomware in real-time. This method is capable of detecting ransomware activity within 100ms, making it suitable for high-speed, real-time detection scenarios. DeepWare has demonstrated a recall of 98.6% and is especially effective against zero-day ransomware families. Dependence on hardware performance counter-supporting platforms restricts its utilization on devices lacking that hardware, making it inelastic compared to other approaches.

BTLS Algorithm: BTLS, which stands for Binary Transformation and Lightweight Signature, was proposed by Wasoye et al. [7] It combines the features of both static and dynamic analysis to enhance the accuracy in the classification of ransomware. The BTLS algorithm significantly reduces false positives while maintaining high classification accuracy by integrating file-based features, such as opcode sequences and binary signatures, with runtime behavior like API calls and network activity. While it does well in terms of accuracy, the computational cost of handling large-scale datasets is a drawback that may not be suitable for resource-constrained environments.

Hybrid Cryptographic Models: Kalphana et al. [8] proposed hybrid cryptographic models that integrate deep learning techniques like AlexNet with cryptographic schemes for enhanced ransomware detection, particularly for Android devices. These models combine DL's ability to extract complex features with cryptographic techniques to secure and classify data. The proposed models achieve an accuracy of 99.89% in detecting Android ransomware. Their effectiveness remains in platforms like Android. They face challenges when attempting implementation across diverse systems.

2.4 Summary of Literature

These works represent a massive leap into ransomware detection, unleashing the power of machine learning and deep learning in this domain but still, several issues remaining, including the preparation of large-scale, diverse, and timely updated ransomware datasets, heavy computational resource costs, and the endless war between ransomware developers and the detection model. Future work should be directed at overcoming these obstacles while exploring new techniques such as adversarial defenses and quantum computing to further strengthen the detection of ransomware attacks.

Technique/Approach	**Key Contributions**	**Limitations**
Opcode Analysis [3]	Early detection through opcode sequence analysis; scalable to real-time environments.	Requires large opcode datasets; computational overhead.
RANSOMNET+ [2]	High precision and recall for cloud environments; combines CNNs and transformers.	Computationally expensive; limited generalizability to non-cloud platforms.
DeepWare [4]	Real-time detection using hardware performance counters; robust against zero-day variants.	Requires platforms with HPC support; limited scalability.
BTLS Algorithm [7]	Integrates static and dynamic features for improved accuracy and reduced false positives.	High computational cost for large-scale environments.
Hybrid Cryptography [8]	Effective for Android ransomware detection with hybrid cryptographic models and deep learning.	Platform-specific; challenges in cross-platform implementation.

Table 1. Summary of Major Ransomware Detection Approaches

Table 1 provides a comprehensive summary of the state of the art in the detection of ransomware using machine learning and deep learning, thus comparing the main strengths and limitations of both. In the further sections of this survey, we will dive deeper into these techniques and provide more detailed insights on how they apply and where there is room for improvement. Each of these techniques has contributed to ransomware detection, though some with better results on different platforms and in various environments. The table further presents trade-offs among the achieved detection accuracy, computational demands, and applicability across various platforms. This becomes very critical during the choice of the right technique for the specified use case. These contributions and limitations need to be realized to give some direction to future research and development of ransomware detection technologies.

Table 2 gives an overview of different hybrid detection models combining static and dynamic analysis with machine learning and deep learning techniques to effectively detect ransomware. The aim is to take advantage of the strengths of several methods and evade the shortcomings of the traditional detection approaches. This table 2 shows how hybrid models are combining several techniques to enhance the efficiency and accuracy of ransomware detection systems. Combining static and dynamic features with machine learning and deep learning methods, these approaches address the limitations of individual detection methods and offer promising solutions for the detection of sophisticated ransomware variants. Challenges such as computational cost, large dataset requirements, and platform dependencies need to be considered when implementing these models in practical environments.

Hybrid Detection Model	Key Features	Advantages/Limitations
Static-Dynamic Integration	Combines static features like file headers with dynamic features like API call sequences. [7]	**Advantages:** Improves accuracy by using a combination of both static and dynamic analysis. **Limitations:** Requires significant computational resources and can be slower than purely static methods.
Opcode-API Hybrid	Combines opcode analysis with runtime behavior observations (API calls and network activity). [3]	**Advantages:** Provides early detection and better understanding of ransomware behavior. **Limitations:** Needs large datasets for training, which can be resource-intensive.
Deep Learning + Static Features	Integrates deep learning techniques (e.g., CNNs) with static analysis features like file signatures. [2]	**Advantages:** Deep learning models improve accuracy by automatically extracting complex features. **Limitations:** High computational cost and significant training data required.
Cryptographic + DL Models	Hybrid models integrating deep learning with encryption-based methods to secure and classify data. [8]	**Advantages:** Highly effective in secure platforms, particularly for Android. **Limitations:** Platform-dependent, not universally applicable.

Table 2. Hybrid Detection Models Combining Static and Dynamic Analysis with ML/DL Techniques

3 Comparative Analysis of Detection Approaches

Table 3 provides a summary of the advantages, disadvantages, and platform-specific applicability of major ransomware detection techniques. Each technique is analyzed based on its strengths and limitations, highlighting the trade-offs involved in their implementation. Opcode Analysis allows for early detection and scalability but requires large opcode datasets, which can be cumbersome to manage. BTLS Algorithm is very accurate and reduces false positives; It is computationally intensive and hence less efficient in a large-scale environment. RANSOMNET+ shows the best performance for cloud-encrypted data, but its applicability is not beyond cloud environments, reducing generalizability. DeepWare is highly effective in real-time detection, resilient to zero-day attacks, though limited to platforms supporting hardware performance counters. Hybrid Cryptography provides high detection accuracy and secure storage; It faces platform-specific challenges that limit its broader implementation. Each technique offers unique strengths, but the associated limitations must be considered when choosing an appropriate solution for specific use cases.

3.1 Platform-Specific Applicability

Table 4 maps each ransomware detection technique to its most suitable platform and provides an overview of its effectiveness metrics. This comparative analysis highlights the applicability and performance of various approaches in distinct environments.

Opcode Analysis shown high detection accuracy of 97.8% with strong generalization across diverse ransomware variants. Scalable and effective for both endpoint and enterprise systems using real-time memory forensics. **BTLS Algorithm**, which has shown a good accuracy of 97% and reduced false positives, applies to general environments. **RANSOMNET+** showed 99.5% Precision and 98.5% recall in cloud environments, making it a go-to tool in cloud ransomware detection. The **Hybrid Cryptography** model, targeted

Technique	Advantages	Disadvantages	Use Case Suitability
Opcode Analysis	Early detection through analysis of low-level code patterns; scalable across different environments	Requires extensive and well-labeled opcode datasets; high computational complexity during feature extraction	Well-suited for offline malware analysis in research and enterprise security labs
BTLS Algorithm	High detection accuracy by integrating static and dynamic features; reduces false positives	Computationally intensive for large-scale datasets; requires monitoring both binary and runtime behavior	Effective in enterprise systems with moderate resources and need for accurate classification
RANSOMNET+	Excellent precision and recall for detecting ransomware in encrypted cloud environments; leverages CNN and transformer architectures	High training and inference cost; limited generalization to on-premise or mobile platforms	Best for cloud service providers and security platforms handling large encrypted data streams
DeepWare	Enables real-time detection within milliseconds using hardware performance counters (HPC); highly resilient against zero-day variants	Limited only to systems supporting HPC; not suitable for lightweight or legacy devices	Ideal for modern enterprise environments with HPC-enabled infrastructure
Hybrid Cryptography	Combines deep learning with cryptographic models to improve ransomware detection, particularly in mobile devices	Platform-specific models; implementation complexity increases in cross-platform systems	Designed for mobile platforms like Android; effective in secure app ecosystems

Table 3. Comparison of well-known ransomware detection techniques

Technique	Platform	Performance Metrics
BTLS Algorithm	Windows, Linux, Enterprise Systems	High accuracy (97%); reduces false positives
RANSOMNET+	Cloud Computing (AWS, Azure, Google Cloud)	Precision: 99.5%, Recall: 98.5%
Hybrid Cryptography	Android	Accuracy: 99.89%; secure storage
DeepWare	Windows, Linux, MacOS	Recall: 98.6%; real-time detection within 100ms
Opcode Analysis	Windows, Linux, Cloud Environments	Accuracy: 97.8%; high scalability and effectiveness across endpoint and enterprise systems

Table 4. Platform-Specific Applicability of Detection Techniques

at Android devices, has an accuracy of 99.89%, with a guarantee of safety in storage, hence making a strong solution for mobile ransomware detection. **DeepWare**, general-platform applicable, provides 98.6% recall and enables real-time detection within 100ms. Hence, it is suitable for environments that have a high demand for quick threat detection. Each method performs well in specific environments and thus is tailored to that particular environment based on its needs.

4 Case Study: Real-World Application of Detection Techniques

The deployment of advanced ransomware detection methods in real-world environments has demonstrated significant potential for mitigating the evolving threat of ransomware attacks. Hybrid models that combine static and dynamic analysis have proven particularly effective in enterprise systems, where diverse ransomware types and operational demands

require versatile and efficient detection solutions. One notable example is the implementation of DeepWare, a deep learning-based model, in live enterprise environments. This approach showcased its capability to detect zero-day ransomware variants in real-time, a critical advancement over traditional detection methods. DeepWare's ability to reduce response time significantly while maintaining high detection accuracy underscores the practical feasibility of deep learning models in dynamic and time-sensitive scenarios [4]. This success is attributed to its reliance on hardware performance counters and deep learning architectures, which enable the model to identify sophisticated attack patterns that evade conventional security measures.

The deployment highlights the relevance of deep learning in combating highly evasive malware. Unlike signature-based or rule-based detection systems, DeepWare adapts to previously unseen ransomware behaviors, leveraging its capacity to generalize from training data to real-world applications. This is especially beneficial in environments where the rapid proliferation of ransomware variants poses a constant challenge to traditional detection methods [5]. The success of DeepWare in live deployments exemplifies the importance of integrating advanced machine learning techniques into security frameworks. Its implementations should address these challenges while exploring additional enhancements, such as integrating adversarial resilience or combining deep learning with quantum-based approaches for even greater effectiveness in detecting and mitigating ransomware attacks.

5 Research Insights from Comparative Analysis

The comparative analysis of ransomware detection techniques highlights key insights that inform both the strengths and limitations of current approaches, as well as opportunities for future advancements in this critical domain:

- **Comprehensive Coverage through Hybrid Models**, Hybrid detection models, which integrate static and dynamic analysis, offer the most extensive coverage by combining multiple perspectives on ransomware behavior. Static analysis enables the identification of known patterns and signatures, while dynamic analysis observes real-time execution behaviors, capturing evasive tactics often employed by ransomware. The complexity of hybrid models introduces challenges in terms of computational overhead and latency, particularly in real-time applications. Optimizing these models to reduce processing time without compromising detection accuracy is a crucial area for further research. Techniques such as model pruning, lightweight architectures, and hardware acceleration could be explored to address these challenges [3].
- **Domain-Specific Excellence with Deep Learning Models**, Deep learning-based methods, such as RANSOMNET+, excel in specialized environments, particularly in cloud systems where they leverage large-scale data and compute capabilities. These models exhibit exceptional precision and recall, often exceeding 99%, making them highly effective in detecting ransomware attacks targeting cloud infrastructure. Nevertheless, their dependency on significant computational resources and their limited generalizability to other environments remain limitations. Expanding the adaptability of such models through domain-transfer learning or hybrid frameworks could enhance their usability across diverse platforms [2].
- **The Role of Enhanced Datasets**, the efficacy of ransomware detection models is highly dependent on the quality and diversity of the datasets used for training and validation. Many existing models are constrained by limited, outdated, or non-representative datasets, which reduce their ability to generalize to novel threats. Future research must focus on curating large-scale, diverse, and frequently updated ran-

somware datasets that encompass a broad range of attack types, platforms, and behavioral patterns. The inclusion of synthetic data generated through adversarial techniques could further enrich these datasets, preparing models to handle emerging threats [9].

– **Strengthening Adversarial Robustness**, with the increasing sophistication of ransomware attacks, adversaries often employ techniques such as obfuscation, polymorphism, and adversarial samples to evade detection. Enhancing the robustness of detection models against such strategies is critical. Future work should focus on integrating adversarial training techniques, wherein models are exposed to adversarial examples during training to improve resilience. Employing ensemble methods or leveraging cryptographic primitives alongside machine learning could further bolster model defenses against adaptive ransomware threats.

While hybrid and deep learning approaches demonstrate considerable promise in ransomware detection, significant challenges remain. Optimizing real-time performance, improving dataset diversity, and enhancing adversarial robustness are pivotal research directions that will drive the next generation of detection methodologies. Addressing these aspects holistically will contribute to more resilient and scalable solutions capable of mitigating the ever-evolving threat landscape.

6 Challenges and Future Research Directions

This section discusses current challenges with ransomware detection models and presents some future research directions that may help to surmount these challenges.

6.1 Challenges in Ransomware Detection

There are still various key challenges in ransomware detection that impede the development of more efficient and accurate detection models. These challenges range from the dynamic nature of ransomware tactics, limitations of datasets, to the increasing demand for real-time detection across diverse environments. Meeting these challenges is crucial for the improvement of ransomware defense systems.

Real-World deployment challenges. Although many detection methods have demonstrated high accuracy in controlled environments, deploying these models in real-world production settings presents significant challenges. Enterprises often rely on legacy systems that are incompatible with modern, high-computation frameworks, making seamless integration difficult. Additionally, detection models trained on carefully curated datasets may struggle to generalize when faced with dynamic, unstructured, and noisy enterprise traffic, leading to potential gaps in threat identification.

Latency is another critical concern, especially in mission-critical systems where even slight delays can have catastrophic consequences, such as financial losses, service disruptions, or security breaches. Moreover, the financial and operational burden of integrating advanced detection mechanisms into existing cybersecurity infrastructures can be a major deterrent. The cost of acquiring, deploying, and maintaining AI-driven security solutions—alongside the complexity of adapting them to an organization's unique IT ecosystem—can slow adoption.

Beyond technical and financial barriers, organizational readiness plays a crucial role in successful implementation. Employees require adequate training to interpret and act upon

AI-generated alerts effectively, preventing both false positives and missed threats. Additionally, regulatory and legal constraints, such as data protection laws, may limit access to sensitive information necessary for model training, further complicating implementation.

These challenges underscore the urgent need for context-aware, scalable, and resource-efficient cybersecurity solutions. Future detection frameworks must not only achieve high accuracy but also balance computational efficiency, interoperability with legacy systems, and compliance with evolving regulatory requirements. A holistic approach—incorporating robust model adaptation, real-time threat intelligence, and human-in-the-loop decision-making—will be essential for ensuring the practical viability of AI-driven security solutions in enterprise environments.

Evolving Evasion Techniques. Ransomware developers constantly innovate and adopt more and more sophisticated evasion techniques, including polymorphism, encryption, and advanced sandbox detection methods. These techniques enable malware to change its code or impersonate legitimate processes, often bypassing traditional and even modern detection systems [9]. Such is the case with most modern ransomware variants, which are polymorphic in nature, meaning they can change their code structure every time they are executed, thus evading signature-based detection systems. To this end, more adaptive detection systems need to be developed which can identify previously unseen variations of ransomware, even if they don't match known patterns [10]. Anti-sandbox techniques such as delaying malicious payloads until they are executed outside the sandbox environment make traditional dynamic analysis ineffective in numerous instances.

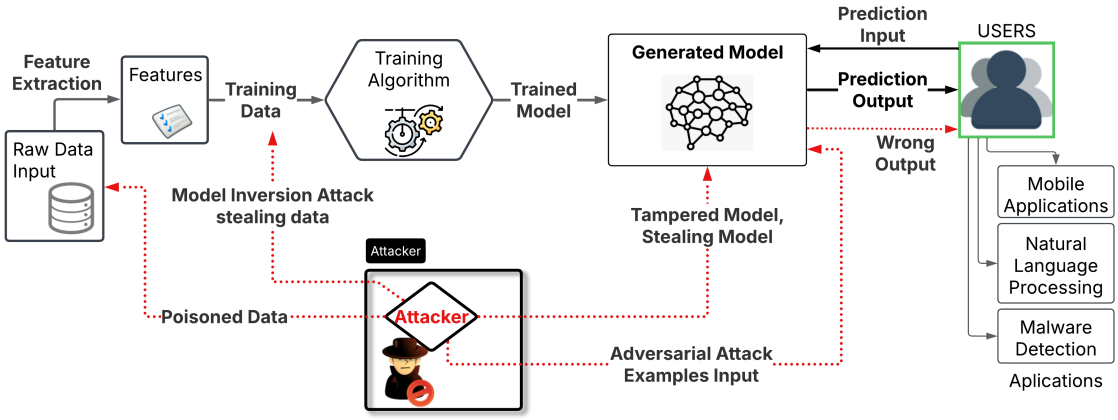

Fig. 1. Illustration of an Adversarial Attack on a Ransomware Detection Model [11]

Adversarial attacks exploit vulnerabilities in machine learning models by deliberately altering input features to mislead their predictions. These attacks can be subtle yet highly effective, as even minor modifications to key attributes—such as opcode sequences in malware detection or API call patterns in behavioral analysis—can cause significant misclassification. Such perturbations are often imperceptible to humans but can drastically shift the model's decision boundaries, exposing weaknesses in its generalization capabilities.

Figure 1 provides a visual representation of this phenomenon, demonstrating how small, targeted alterations in input data can lead to erroneous outputs. This underscores the critical need for developing adversarially robust models capable of resisting such manipulations. Enhancing resilience against adversarial attacks requires incorporating advanced defense

mechanisms, such as adversarial training, feature-space regularization, and anomaly detection techniques. By fortifying models against these deceptive tactics, researchers and practitioners can improve the reliability and security of machine learning systems in high-stakes applications.

Dataset Limitations The lack of diverse and comprehensive datasets remains one of the primary challenges in developing machine learning- and deep-learning-based ransomware detection models. Most available datasets are either outdated or with limited variety, making it difficult for models to detect emerging ransomware variants [5]. Apart from this, most ransomware detection models are trained using datasets that consist predominantly of a few types of malware, making them behave poorly when faced with new attacks. There is an urgent need to create more robust and diversified datasets that involve a larger area of ransomware behaviors due to continuous developments within the genre. Ideally, these data sets should be broad in spectrum and constantly updated with new samples to train models capable of generalizing throughout the malware spectrum [7].

Real-Time Detection Constraints Real-time detection is another important challenge in ransomware detection. To be effective, detection systems have to be highly accurate and computationally efficient. The strength of most modern deep learning models involves immense computational overheads and is not suitable for a resource-constrained environment. In particular, such systems with low hardware capability or deployed on embedded devices often run a heavy neural network with significant latency. For this challenge, research has focused on constructing lightweight models without sacrificing accuracy but without extreme computational overhead [12]. Such performance versus efficiency trade-offs continue to be an important direction of research.

Cross-Platform Challenges Often, a ransomware detection model is designed with some specific platform in mind, such as Windows or Android. This dependency on the platform reduces the applicability of these detection systems in heterogeneous environments where multiple platforms are used. example, although such systems as DeepWare perform extremely well on specific platforms, their generalization to numerous operating systems or hardware configurations has issues of scalability [4]. With the evolution of ransomware towards targeting multiple diverse platforms, the demand is rapidly growing for cross-platform ransomware detection systems. The research efforts should be focused on the development of platform-agnostic detection models using features that can be applied to any environment while ensuring robustness and scalability [13].

6.2 Future Research Directions

This section describes into the future direction for ransomware detection should focus on enhancing model robustness through adversarial training and leveraging advanced technologies like quantum computing.

Adversarial Defenses To enhance robustness in ransomware detection systems, the use of adversarial machine learning techniques has vast potential. For instance, adversarial training-exposing the detection models to modified ransomware variants during training-shows increased resistance against evasion attacks [14]. This approach aims to make the detection systems more resistant to sophisticated attacks by preparing them for new variants that might try to deceive the system using adversarial methods. Since ransomware

developers increasingly use adversarial strategies, research into adversarial defenses is important to maintain the efficacy of ML-based detection systems [9].

Quantum Computing Quantum computing holds revolutionary potential for ransomware detection, especially in handling large datasets and real-time processing demands. Quantum machine learning algorithms, including quantum support vector machines, have the ability to process large volumes of data in an exponentially quicker way compared to traditional methods, which could be a game-changer for real-time ransomware detection [10]. Their application to ransomware detection is still purely theoretical, but with significant promise to overcome current computational challenges in the field. Future research should explore the integration of quantum computing with traditional ML models to accelerate malware detection and response times.

Hybrid Models with Federated Learning Integrating federated learning with hybrid detection models is another promising direction. Federated learning allows for the training of ML models across decentralized devices, where the data remains on the local device, thus addressing privacy concerns. These decentralized devices' aggregated insights can be utilized to train a global model without necessarily having to exchange the raw data. This may significantly improve the scalability and privacy of ransomware detection systems, especially in environments where data privacy is a major concern [8]. In addition, such a decentralized approach might offer a robust mechanism for the detection of ransomware in environments where traditional centralized systems might result in a single point of failure.

Federated learning facilitates decentralized model training across multiple devices while ensuring that raw data remains local, thereby preserving user privacy. This paradigm is particularly advantageous in highly regulated sectors such as healthcare and finance, where stringent legal and ethical considerations restrict the sharing and centralization of sensitive data. By enabling collaborative model development without requiring direct access to proprietary datasets, federated learning fosters the creation of robust and generalizable machine learning models while mitigating privacy risks associated with conventional centralized approaches. Potential federated learning presents several critical challenges that must be addressed for effective implementation. One of the primary concerns is the issue of non-independent and identically distributed (non-IID) data, wherein data distributions vary significantly across participating devices or institutions. This heterogeneity can lead to inconsistencies in model updates, resulting in suboptimal performance and reduced convergence rates. Federated learning introduces considerable communication overhead, as frequent model updates between distributed nodes and the central aggregator require substantial bandwidth and synchronization efforts and resource-constrained edge devices often lack the computational power necessary for efficiently training complex models, further complicating large-scale deployment.

Some challenges, the integration of federated learning with ransomware detection models offers a promising avenue for enhancing security and privacy in distributed computing environments. By leveraging federated learning, ransomware detection systems can continuously adapt to emerging attack vectors across multiple platforms without exposing sensitive organizational or user data. This decentralized approach not only strengthens the robustness and adaptability of threat detection mechanisms but also aligns with regulatory frameworks governing data privacy and security. As a result, federated learning emerges as a compelling solution for developing privacy-preserving, scalable, and resilient cybersecurity models in an increasingly interconnected digital landscape.

6.3 Expansion of Datasets

One of the most important directions for future research is the expansion of datasets used to train ransomware detection models. Current datasets are often small and limited in variety, which hinders the development of generalized models that can detect diverse strains of ransomware. The researchers need to develop comprehensive datasets that contain not only samples from commonly seen ransomware families but also new and emerging variants, especially from new platforms, such as IoT devices and cloud environments [8]. These will help the models to adjust with the evolution of ransomware and stay effective against any future malware.

Cross-Platform Detection Frameworks Since ransomware threats continue to evolve, the immediate requirement would be towards unified detection frameworks working across multi-platforms from desktop operating systems like Windows and Linux to mobile platforms such as Android and iOS. The nature of these detection frameworks needs to be adaptable with unique properties on each platform but based on similar techniques so that consistency and reliability may be maintained. Cross-platform solutions in the research should aim to build detection models that can generalize to various environments without sacrificing performance and scalability [5].

6.4 Emerging Trends

Recent advances in ransomware detection now embed more threat intelligence platforms, providing improved predictive analysis and real-time intelligence into the latest tactics of ransomware [15]. Besides that, generative models like Variational Autoencoders VAE are also gaining traction for zero-day detection and showing promising results toward detecting unseen variants of ransomware [10].

Threat Intelligence Integration One such trend is integrating ransomware detection systems with threat intelligence platforms. This can have potentially considerable effects on enhancing predictability. Feeding in real-time threat intelligence information into the detection models keeps the systems proactive in finding an emerging variant of ransomware before it spreads all over. This can be realized based on external data from threat reports, malware behavior analysis, and global security trends that proactively provide early warning for attacks against ransomware [15].

6.5 Zero-Day Detection

Zero-day detection is considered one of the most important research areas in this field, as ransomware developers continue to take advantage of previously unknown vulnerabilities. Traditional signature-based methods are ineffective against zero-day threats, making anomaly detection techniques, such as Variational Autoencoders , crucial for identifying malicious behavior without prior knowledge of the malware [10]. This research direction is going to enhance the detection systems' capability to identify ransomware variants that have not been seen by any signature-based system.

6.6 Opportunities for Collaboration

Collaboration between academia, industry, and government agencies is vital for accelerating advancements in ransomware detection and the development of more robust cybersecurity defenses. By forming strategic partnerships, these sectors can leverage their

unique strengths to create shared datasets, develop standardized evaluation benchmarks, and drive innovation in both detection methodologies and defense strategies. Academia can contribute cutting-edge research and theoretical insights, while industry can provide practical, real-world data and the necessary infrastructure for large-scale implementation. Government agencies, on the other hand, can offer regulatory guidance and support in terms of public policy, as well as facilitate collaboration across various sectors of society.

Such collaborations are crucial for the establishment of best practices in ransomware mitigation. When these entities work together, they can develop coordinated, comprehensive approaches to address the complexities of ransomware threats. This could involve creating common frameworks for evaluating the effectiveness of detection systems, ensuring that models are tested and validated across a broad range of attack scenarios and environments. Joint efforts in sharing threat intelligence and research findings could lead to faster identification of emerging attack vectors and more timely updates to defense systems. Ultimately, a unified approach to ransomware detection and prevention will not only improve the technical capabilities of defense systems but also promote a more holistic and integrated cybersecurity strategy across different sectors [9].

6.7 Summary of Challenges and Opportunities

Table 5 summarizes the key challenges and future research directions in ransomware detection, providing a roadmap for researchers to navigate the evolving landscape of ransomware defense.

Challenges	**Future-Opportunities**
Evolving evasion techniques	Adversarial defenses and robust ML models
Dataset limitations	Expansion of diverse and comprehensive datasets
Real-time detection constraints	Quantum computing and federated learning
Cross-platform challenges	Unified frameworks for multi-platform detection

Table 5. Challenges and Future Research Opportunities in Ransomware Detection

The table 5 summarizes some of the key challenges in ransomware detection, with related future opportunities to surmount those challenges. First, “Evolving evasion techniques” demands more sophisticated detection methods that can resist polymorphic and other evasion strategies, whereas “adversarial defenses” and “robust machine learning (ML) models” may help in countering these evasions as suggested by future research [9]. Another major obstacle to accurate model development is “dataset limitations, where the diversity and size of the data that are available are insufficient”, with a promising opportunity lying in “comprehensive dataset expansion to wider ranges of ransomware variants”[7]. The real-time detection remains a challenge, since most techniques are with high computational cost and latency; leveraging emerging technologies such as “quantum computing” and “federated learning” can enhance significantly the speed and scalability of the detection models, thus allowing for real-time performance with much more efficiency [10]. Finally, “cross-platform challenges” are related to the difficulty of developing detection systems that work on different operating systems and environments. The development of “unified frameworks” which can be adapted to different platforms will be of great importance to enhance versatility and effectiveness in ransomware detection systems [8].

7 Conclusion

Ransomware remains one of the most significant threats to global cybersecurity, constantly evolving to bypass traditional detection mechanisms. This survey has underscored the critical role of machine learning and deep learning in mitigating these advanced threats. We have examined a variety of contemporary detection methodologies, including opcode analysis, hybrid cryptographic models, and cutting-edge deep learning frameworks such as RANSOMNET+ and DeepWare. Our comparative analysis highlights the strengths and limitations of these approaches across various platforms, emphasizing the superior efficacy of hybrid detection techniques in achieving comprehensive threat mitigation. These advances, several formidable challenges persist. The increasing sophistication of ransomware evasion techniques, coupled with the limitations of existing datasets, hampers the development of more resilient detection systems. Addressing these challenges requires continuous innovation, rigorous evaluation, and refinement of machine learning-driven cybersecurity solutions. Future research should prioritize the development of scalable, platform-agnostic, and adaptive models capable of countering the rapidly evolving nature of ransomware. Emerging technologies such as adversarial defense mechanisms, quantum computing, and federated learning present promising opportunities to enhance the accuracy and efficiency of ransomware detection frameworks. In addition, interdisciplinary collaboration between academia, industry, and government agencies is essential to overcome these challenges. Such partnerships can facilitate the creation of shared datasets, the establishment of standardized evaluation metrics, and the development of innovative defense strategies tailored to the evolving ransomware landscape. Using collective expertise and resources, the cybersecurity community can devise more robust countermeasures to address emerging threats. This paper establishes a foundation for future research and practical advances in ransomware detection. Continued efforts in this domain will contribute to the development of more resilient cybersecurity frameworks, ensuring timely and effective mitigation strategies against ransomware attacks. By fostering sustained innovation and cross-sector cooperation, the cybersecurity field can strengthen global defenses against this ever-evolving threat, thereby fostering a safer and more secure digital environment.

8 Acknowledgment

This research was conducted as part of the KISON group, supported by the Universitat Oberta de Catalunya. The authors express their gratitude to the KISON team for their valuable contributions and support throughout this study. My work was supported by the KISON research group at UOC under its ongoing research efforts in cybersecurity and networking.

The authors acknowledge the funding obtained by the grants: Detection of fake newS on SocIal MedIa pLAtfoRms (DISSIMILAR) from the EIG CONCERT-Japan (PCI2020-120689-2, Government of Spain), the “SECURING” project (PID2021-125962OB-C31) funded by the Ministerio de Ciencia e Innovación, la Agencia Estatal de Investigación and the European Regional Development Fund (ERDF), as well as the ARTEMISA International Chair of Cybersecurity (C057/23) and the DANGER Strategic Project of Cybersecurity (C062/23), both funded by the Spanish National Institute of Cybersecurity through the European Union – NextGenerationEU and the Recovery, Transformation and Resilience Plan.

References

1. Amjad Alraizza and Abdulmohsen Algarni. Ransomware detection using machine learning: A survey. *Journal of Cybersecurity*, 16:118–130, 2021.
2. Amardeep Singh, Zohaib Mushtaq, Hamad Ali Abosaq, Salim Nasar Faraj Mursal, Muhammad Irfan, and Grzegorz Nowakowski. Enhancing ransomware attack detection using transfer learning and deep learning ensemble models on cloud-encrypted data. *Electronics*, 12(18):3899, 2023.
3. Benjamin Pesem, James Fairweather, and Thomas Pennington. Opcode memory analysis: A data-centric machine learning framework for early detection and attribution of ransomware. *IEEE Transactions on Computers*, 72(3):600–613, 2024.
4. Gaddisa Olani Ganfure, Chun-Feng Wu, Yuan-Hao Chang, and Wei-Kuan Shih. Deepware: Imaging performance counters with deep learning to detect ransomware. *IEEE Transactions on Computers*, 72(3):600–613, 2022.
5. Jinsoo Hwang, Jeankyung Kim, Seunghwan Lee, and Kichang Kim. Two-stage ransomware detection using dynamic analysis and machine learning techniques. *Wireless Personal Communications*, 112(4):2597–2609, 2020.
6. Umara Urooj, Bander Ali Saleh Al-rimy, Anazida Zainal, Fuad A Ghaleb, and Murad A Rassam. Ransomware detection using the dynamic analysis and machine learning: A survey and research directions. *Applied Sciences*, 12(1):172, 2021.
7. Samuel Wasoye, Michael Stevens, Christopher Morgan, David Hughes, and James Walker. Ransomware classification using btls algorithm and machine learning approaches. 2024.
8. KR Kalphana, S Aanjankumar, M Surya, MS Ramadevi, KR Ramela, T Anitha, N Nagaprasad, and Ramaswamy Krishnaraj. Prediction of android ransomware with deep learning model using hybrid cryptography. *Scientific Reports*, 14(1):22351, 2024.
9. Craig Beaman, Ashley Barkworth, Toluwalope David Akande, Saqib Hakak, and Muhammad Khurram Khan. Ransomware: Recent advances, analysis, challenges and future research directions. *Computers & Security*, 111:102490, 2021.
10. Thomas Lowev, Charlotte Fisher, and James Collins. Advanced ransomware detection and classification via semantic analysis of memory opcode patterns. 2024.
11. Miguel Hernández. Adversarial machine learning: A beginner's guide to adversarial attacks and defenses. `https://hackernoon.com/adversarial-machine-learning-a-beginners-guide-to-adversarial-attacks-and-defenses`, 2022. Accessed: 2025-03-24.
12. Chia-Ming Hsu, Chia-Cheng Yang, Han-Hsuan Cheng, Paul E Setiasabda, and Jenq-Shiou Leu. Enhancing file entropy analysis to improve machine learning detection rate of ransomware. *IEEE Access*, 9:138345–138351, 2021.
13. Amnah Albin Ahmed, Afrah Shaahid, Fatima Alnasser, Shahad Alfaddagh, Shadha Binagag, and Deemah Alqahtani. Android ransomware detection using supervised machine learning techniques based on traffic analysis. *Sensors*, 24(1):189, 2023.
14. Michael Argene, Clara Ravenscroft, and Ivy Kingswell. Ransomware detection via cosine similarity-based machine learning on bytecode representations. 2024.
15. Daryle Smith, Sajad Khorsandroo, and Kaushik Roy. Machine learning algorithms and frameworks in ransomware detection. *IEEE Access*, 10:117597–117610, 2022.

Authors

Muhammad Junaid Iqbal received a Bachelor's degree in Information Technology from GCU, Faisalabad, Pakistan, and later completed his Master's degree in Computer Science from FAST-NU, Islamabad, Pakistan. He is currently pursuing a PhD in Computer Science at the Universitat Oberta de Catalunya (UOC). His research interests include information security, cryptography, cybersecurity, and the application of artificial intelligence in cybersecurity.

Jordi Serra-Ruiz received a Master's degree in Industrial Computing from UAB University, Barcelona. He obtained his PhD in Computer Engineering from the Universitat Oberta de Catalunya (UOC), Barcelona. His research interests include information security, cryptography, cybersecurity, and the application of artificial intelligence in cybersecurity.