IJEMD-CSAI, Special Issue (2026) https://doi.org/10.54938/ijemdcsai.2026.04.2.659

IJEMD

International Journal of Emerging Multidisciplinaries: Computer Science and Artificial Intelligence

Research Paper

Journal Homepage: www.ojs.ijemd.com
ISSN (print): 2791-0164 ISSN (online): 2957-5036

Ransomware Defence Systems

Cheah Zhen Yang [1], Joshua Angelo Kana [1], Arvind Skanda Dharmendran [1], Muhamad Haziq Bin Jasni [1], Dinesh A/L S Rajandran [1], Ung Qi Hang [1], Wan Efdlin Eirfan [1], Siva Raja Sindiramutty [1]

1. School of Computer Science, Taylor's University, Subang Jaya, 47500, Selangor Malaysia

Abstract

Ransomware has quickly become one of the newest and most harmful types of cyberattacks. Traditional security programs like antivirus programs, firewalls and signature-based Endpoint Detection and Response (EDR) programs are no longer good enough, particularly in dealing with the zero-day attacks and the new types of ransomwares that evolve their behavior to evade detection. This creates a major threat to organizations, such as operational disruptions, loss of money and irreparable data breach. To overcome these shortcomings, the Hybrid Ransomware Defense Framework (HRDF) is presented in this report, a multi-layered and dynamic security framework that will provide comprehensive protection at every phase of a ransomware attack. HRDF combines AI-based anomaly detection, deception-based honeypot validation, Zero Trust security policy, and automated responses in a single, unified defense platform. Its detection layer uses a dual-model machine learning approach, which incorporates both Random Forest classification and LSTM-based behavioral analysis as its means of detection, which allows the system to detect both structural anomalies and time-based attack patterns in real time. The automation that is inherent in HRDF significantly lowers the Mean Time to Detect (MTTD) as well as the Mean Time to Respond (MTTR), hence quick containment and recovery. In general, HRDF offers end-to-end ransomware protection that is beyond conventional approaches, and it includes proactive detection, trusted prevention, and automated recovery that is smooth enough to fit today's advanced threat environment.

Keywords: Ransomware Defense Framework, AI-Based Anomaly Detection, Hybrid Machine Learning (Random Forest & LSTM), Zero Trust Architecture, Automated Incident Response and Recovery

Email Addresses : 0361648@sd.taylors.edu.my, 0362637@sd.taylors.edu.my, 0362384@sd.taylors.edu.my, muhamadhaziq.jasni@sd.taylors.edu.my, dinesh.srajandran@sd.taylors.edu.my, qiho.ng@sd.taylors.edu.my, 0379136@sd.taylors.edu.my, magan.shiva91@gmail.com

1.0 Introduction and Problem identification

1.1 Introduction to Ransomware

In this world that we live in where technology has woven itself so tightly into almost every aspect of our life's and benefitted us greatly, it can often also pose as threat to us too if not properly used and controlled. Among all of these threats there exists Ransomware, attackers use it to prey upon individuals, organizations and even large corporations. To begin we must first establish a baseline understanding on what "Ransomware" is. Ransomware is a form of malware that takes the targets data and information hostage by encrypting it, the attacker will threaten to keep the data encrypted or destroy it unless the ransom is paid by the victim (Kosinski, n.d.).According to a study, 98% of ransomware attacks demand Cryptocurrency such as Bitcoin as the requested payment form by the attacker. This is due to the fact that the payment can easily be confirmed since the blockchain is public. It also adds a layer to the attacker's anonymity as none of the attacker's information can be found when laundered through various wallets (Amado, 2021; Riskhan et al., 2025; Humayun et.al., 2022).

1.2 Evolution of Ransomware

By looking and learning from the key moments in history and evolution of ransomware, we will be better able to understand trends and discover patterns that may help us in the future. The first officially documented case of a ransomware was one called AIDS Trojan in the year 1989. The attack was mostly unsuccessful as it only encrypted the names of the files and not the entirety of it. It then prompted the victim to pay $189 to PC Cyborg Corporation to get a "repair" tool. Another flaw in the attack was that the decryption tool was already inside of the Trojan (Keon et al., 2025).

An idea was put forth in 1992 by Sebastiaan von Solms and David Naccache, utilizing anonymous digital cash systems, one could safely collect ransom from kidnapping victims. In a scenario dubbed the von Solms-Naccache scenario, a newspaper publication was used. So that payment could be demanded without revealing their identity through coded messages in newspapers. This is important as it sets up the foundation for how ransomware attackers will use this idea to request payment from victims moving forward.

2006 marked the emergence of advanced RSA encryption. A few notable ones were Arhiveus, Cryzip and Gpcode. A variant of Gpcode, Gpcode.AG, was detected in June of 2006, it used a 660-bit RSA public key. In June, a different variant called Gpcode.AK was found. This time it boasted a 1024-bit RSA key. At the time it was assumed to be computationally impossible to crack it. In the 2010s, Bitcoin among other cryptocurrencies began gaining popularity mainstream. As previously stated, the decentralized and distributed nature of cryptocurrency made this the perfect match for ransomware payment. As it was easy to obfuscate and evade any investigation attempts by law enforcement (Arctic Wolf, 2024; Kiyani et al., 2024; Fatiam et.al., 2020).

Ransomware-as-a-service (RaaS) got its big break in 2012 via Reveton ransomware. It would impersonate local law enforcement and threaten to arrest or detain victims if they did not pay a fee. The people behind Reveton also sold the malware to third parties a as a service. This made the malware spread even faster at an unprecedented rate (Arctic Wolf, 2024; Krishnan et al., 2021). The year after that, a ransomware was spread via a botnet and social engineering. It was called CryptoLocker. Its main way of spreading was via email attachments that were exacerbated by the botnet. According to

ZDNet, a business technology news website, CryptoLocker gained an approximate amount of $27 million dollars from the victims of its malware

In 2016, a malware called Petya was the first to ever overwrite the master boot record and encrypt the master file table. It was able to hijack the system at its core and subvert its start-up process, all while being more destructive and faster. The following year, a variant of it called NotPetya brought many Ukrainian infrastructure and systems to a standstill. It started when attackers compromised an accounting software called M.E.Doc. Since it was used by many businesses in various sectors, all of them were compromised. Even neighbouring countries and allies of Ukraine were affected, such as the UK, France and US. Many experts have concluded that this attack was carried out by Russia who had an ongoing conflict with them

2017, the same year as NotPetya, another ransomware attack took headlines across the world. This time it was WannaCry, it was the biggest ransomware attack in history as it spread across 150 countries. It originated in Asia via phishing and spread rapidly due to taking advantage of a Microsoft vulnerability known as Eternal Blue (Arctic Wolf, 2024; Linqiang et al., 2024).

Starting from 2018, attackers started implementing data exfiltration, where the malware would exfiltrate and steal the victim's data such as personal information, files, payment info, etc. In recent years attackers have become more specialized by scouting and picking high value targets to attack for profit. This has been called Big Game Hunting, the targets of these attacks are usually high net worth individuals or organizations

1.3 Limitations of Current Security Solutions

It is very evident from this timeline that ransomware is rapidly evolving and getting more and more dangerous. As such, when designing and proposing a Ransomware defence system, a lot of consideration must be made at the current landscape of cybersecurity and ransomware. Many of the readily available solutions are lacking in certain areas. There are a few reasons as to why the solutions currently available on the market are inadequate. Firstly, most of them rely heavily on signature-based detection. This put them at a high risk of a zero-day ransomware as lacks means to detect and identify new strains. Newer forms of malware also use fileless techniques by running off the memory or by using PowerShell and other system tools to leave no signatures for the defence system to catch on (Baker, 2024; Sanjaya et al., 2025). Secondly, endpoint resilience is severely lacking with these new solutions. Even with the latest antivirus and EDR tools, endpoints such as PCs, laptops and workstations are often still the first point of infection. Furthermore, once modern malware executes, it can usually disable or bypass security software and prevent any basic countermeasures or recovery points. Lastly, there is usually a disconnect or rift between the detection and response parts of these solutions. From the point of detection to response it takes too long and allows the malware to inflict damage and steal data before containment and further actions (Field Effect, 2025; Sindiramutty et al., 2024; Sobia et.al., 2025).

1.4 Real-World Impact

It is crucial that everyone, especially business owners and organizations understand the damage a ransomware attack can do. To name a few recent examples such as the one in 2021, called the Colonial Pipeline Company Attack that shutdown the system that is responsible for 45% of the fuel used on the East Coast in US. It caused major disruptions and price spike to fuel prices (William Turton, 2021;

Sindiramutty et al., 2024b). Another one such as the MOVEit Transfer Breach that took place in 2023. It took advantage of a zero-day vulnerability and impacted millions of people worldwide. It affected multiple chains of supplies and infrastructures. This goes to show that even huge organizations are not immune to these attacks, if anything they are more vulnerable since they are the prime targets (Page, 2023). It is up to the people to take the effort to secure themselves and their businesses and organizations with the best solution that can address all these core issues.

1.5 Problem Statement

Ransomware has evolved well beyond the traditional signature-based detection, fileless execution, double extortion, and zero-day vulnerabilities, and modern security tools are not able to detect it. Most organizations continue to use old solutions that are unable to identify new variants at the initial stages, prevent the further movement of the virus, and restore the systems as soon as they are affected. This causes severely important security lapses when ransomware can encrypt data, steal information, and shut down operations before any response measures are taken into action. Thus, it is urgent to have a fast, adaptive and automated ransomware defense system able to identify threats in time, authenticate them, and act in real time.

These issues are a clear indication of why the traditional defenses are not enough anymore. The increasing speed, complexity and scale of ransomware attacks require a multi-layered approach that is more than simple detection. This sets the foundation for the Hybrid Ransomware Defense Framework (HRDF), incorporating behavioral AI detection, honeypot validation, Zero Trust controls, and automated recovery to mitigate the vulnerabilities of the modern cybersecurity solutions.

2.0 Literature Review & Gap Analysis

2.1 Literature Review

Recent studies have talked about different approaches in recognizing anomalous behavior to prevent intruders into systems. Feng et al. (2024) focused on creating an anomaly detection model combining coarse-grained and fine-grained detection modules to emphasize the recall rate of anomaly detection and increase precision rate of anomaly detection respectively. The proposed model used oversampling and under sampling techniques, training on historical based data models and validates future behavior data together with publicly available user anomaly detection dataset CERT. Although Feng et al. proved the model able to improve accuracy and strongness of internal threat detection, Feng et al admits that his study uses metrics under the same condition to only find detected users and not further into the degree of abnormality behaviors. Feng et al. said the differences in anomaly detection ability for various users did not appear to be significantly correlated with sample size. This implies that various users exhibit unique patterns of behavior and distributions of normal and anomalous behavior samples, which leads to the conclusion that the ideal configuration for the quantity of training data varies and must be tailored to each user's unique needs and circumstances.

Folino et al. (2023) tried to tackle the limitations of using a single classification algorithm by providing an elastic stack-based framework for processing and storing user data, creating a group of classifiers to categorize user behavior, and using this categorization to effectively identify abnormalities in user behavior. In actual use, the system employs a distributed evolutionary algorithm for user classification based on digital footprints collected from several logs and the high-performance architecture offered by ELK, operating on top of a Kubernetes-based platform. Furthermore, the framework allows for the individuation of user behavior abnormalities as a new resulting task. Results show that the model can effectively manage missing and unbalanced data sources. One of the limitations to the study includes the semi-supervised approach which limits the model’s ability to learn from actual anomalies hence reducing detection accuracy in novel attack scenarios. Another limitation is the generalizability of the framework could be affected due to assuming the training dataset is complete while real-world data always contains missing values. These studies are particularly relevant to ransomware defense, where early detection of anomalous user behavior can prevent full-scale compromise. Behavioral analytics and ensemble models offer promising tools to identify such threats before payload execution.

Park & Dagher (2025) uses Game Theory, specifically The Bayesian Stackelberg Game to examine the tactical exchanges between logical decision-makers. The model alone won’t be sufficient hence they combined it with a Mixed-Integer Linear Programming (MILP) paradigm to utilize strategically placed honeypots and dynamic belief updates to lead attackers to disclose their intentions early in hopes of enhancing the network's resistance to sophisticated and changing cyberthreats by improving the defender's capacity to predict and disrupt attacks. The study concludes this method not only outperforms conventional baselines like random or greedy honeypot placement, but it also scales well to networks with up to 500 nodes and more than 1,500 edges in run times of about a second. Some of the limitations that they faced were using an approach that did not picture rapid adaptation and no account taken into zero-day exploits, partial observability, or heterogeneous security teams.

Chacon et al. (2020)& attempts to combat APT attacks by using honeypot method due to anomaly-based intrusion being ineffective by providing a high percentage of false negatives, false positives, or

both. Chacon et al uses honeytokens in a self-contained system to prove that it is possible to identify and categorize incursions based on the degree of human reasoning and interaction needed to carry them out or put them up. The experiment for the study was set up to investigate whether the deduction can be done on influence of human reasoning and interactions using honeytokens in honeypots within a perimeter and whether the arrangement of honeytokens can determine the attack level intent. Methods include a few honeytoken techniques, a honeypot set up in a company virtual private infrastructure within a virtual subnet and a couple of pentesters to act as APT attackers. The evaluation of the whole experiment is measured using the severity of the intrusion detected, which are low priority that shows any level of network detected with the honeypot or the application of automated reconnaissance methods that require little human input, medium priority where the reconnaissance's results were utilized to try a breach and advance the intrusion and high priority where after a human actor found honeytokens, more logic was used to increase the efforts, which resulted in a breach. Following positive results and answers to both the questions of the experiment, Chacon et al. agreed that early signs of APT can be detected using this type of deception. One limitation that could be pointed out is that the experiment was set up according to the authors' belief that they were mimicking a real APT actor even though the design shows differences in many aspects from the behavior of a real APT actor.

Seetharam kakaraparthi et al. (2024) took advantage of the decentralized nature and cryptographic principles of blockchain to deploy honeypots in addition to integrating smart contracts which allows for automatic reactions to threats that are discovered, simplifying incident response procedures and cutting down on mitigation time. The idea encourages safe sharing of useful information and compromise indicators (IOCs) by facilitating peer organizations' smooth exchange of threat intelligence, harnessing collaboration and collective defense against cyberattacks. Experiment results were evaluated through performance metrics and measures such as response time, throughput, latency and resource utilization. Then, the system is put through simulated attack scenarios and real-world threat feeds which includes different types of attacks to learn the system's detection capabilities. Stress tests were also done to measure the scalability of the system. Results showed encouraging outcomes, indicating the system's capacity to accurately and efficiently identify and address a range of cyberthreats. Security evaluations emphasized steps made to reduce vulnerabilities and defend against attacks, while performance testing demonstrated the system's scalability and resilience under various workloads. A useful context was supplied by a comparative analysis against baseline benchmarks and current IDPS solutions, highlighting the special benefits of blockchain integration in bolstering cybersecurity defenses. However, the system tested was modeled on general cyber threats instead of ransomware focused ones and the usage of smart contract logic may be limited in handling more complex scenarios.

A study on AI based IDS systems was done by Ajibuwa, Hamdaoui, and Yavuz (2023) that was a survey on AI/ML-Driven Intrusion and Misbehavior Detection in Networked Autonomous Systems. The survey consisted of 36 recent papers which Ajibuwa et al. reviewed to analyze baseline feasibility metrics that were established by Ajibuwa et al. and to identify systematic gaps that limit practical deployment. The taxonomy and scope of this paper is divided into 1) Covered systems, 2) Detection Strategies, 3) Architectures, 4) ML families used, and 5) Evaluation practices.

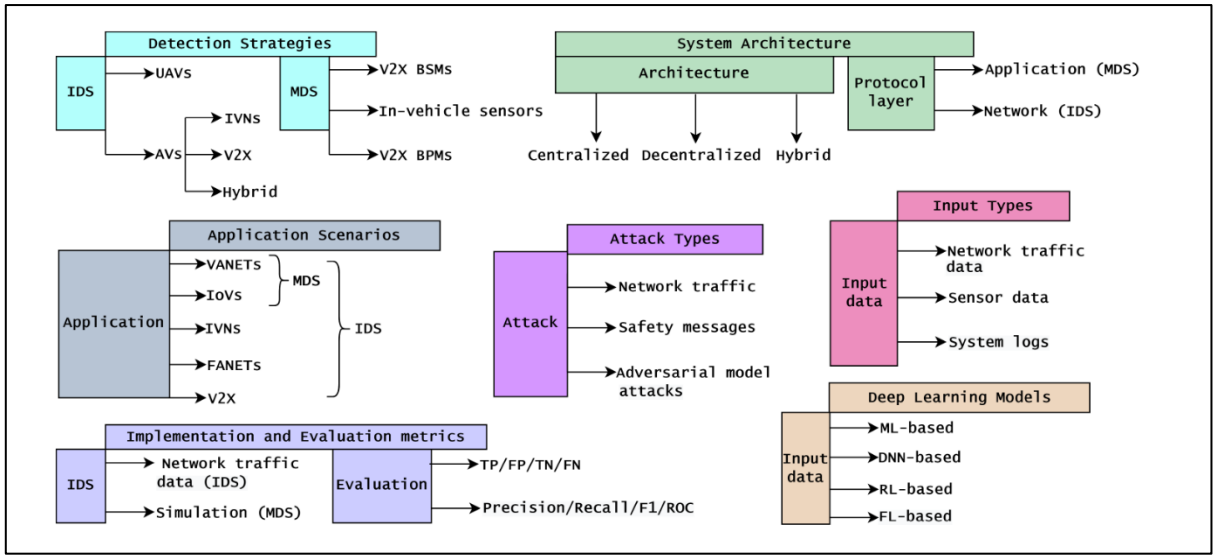

Figure 1 A consolidated overview of IDS/MDS detection strategies and system architectures used in autonomous and connected systems

Figure 1 is an overview of the significant parts incorporated in intrusion detection systems (IDS) and misbehaviour detection systems (MDS). The figure classifies detection strategies into UAV-based, in-vehicle, V2X, and hybrid monitoring strategies; system architecture (centralised, decentralised and hybrid); protocol layers (network or application). It also outlines typical application cases such as VANETs, IoVs, IVNs, FANETs and V2X systems. Types of input data (network traffic, sensor data, system logs) can be used with various deep-learning models, such as ML, DNN, RL, and FL-based. Some of the performance measures that are popularly used are in the evaluation section and include TP/FP/TN/FN, Precision, Recall, F1-score, and ROC curves.

It is reported that the papers that were reviewed left out important metrics that should be considered when deciding whether an IDS is ready to be implemented to handle real life attacks. Thorough analysis helped the authors decide the baseline metrics that needed to be reported to ensure their capabilities in the real world. The baseline metrics are as follows:

Baseline Metrics		
Attack-detection	Hardware & Deployment	Secondary
Accuracy	CPU/processing cost	Price
False Positive Rate (FPR)	Memory footprint	System Complexity
False Negative Rate (FNR)	Bandwidth/Latency impact	Scalability
Detection Time	Energy consumption	Forensics
Training Time	Ease of deployment	Recovery
MTBF	Placement Strategy	
Model Update Latency		
Dataset relevancy		
Privacy		

Figure 2: A consolidated summary of the baseline feasibility metrics identified by Ajibuwa et al. (2023)

Figure 2 is a summary of the baseline metrics suggested by Ajibuwa, Hamdaoui, and Yavuz (2023) to assess the viability of the IDS/MDS solutions to be implemented in real life. The measures fall into three dimensions:

Attack-Detection Metrics: they are accuracy, the number of false positive (FPR), the number of false negative (FNR), the latency of model update, relevance to the dataset, and privacy. Hardware and Deployment Metrics e.g. CPU and cost of processing, memory footprint, bandwidth/latency impact, energy consumption, easy to deploy, and placement strategy. Secondary Metrics such as system complexity, scalability, forensic readiness, recovery capability and price. All these measures indicate the ability of an IDS/MDS to be used in practice subject to real-life limitations, such as cost and performance, scalability and operational feasibility.

There were also limitations in this study that must be taken account into. Firstly, the review covers a curated set of recent papers rather than the entire literature, which can introduce selection bias and may omit relevant studies, datasets or industry reports outside the chosen sample. Then, conclusions depend on how original papers reported metrics. When those papers omitted or reported metrics inconsistently, the survey's assessment of feasibility is limited by incomplete or non-uniform source data. Finally, the

rapid advances in ML model, edge hardware, deployment patterns, and standards mean some finding or recommended metrics may become dated quickly and may not reflect the very latest techniques, datasets, or OEM practices.

2.2 Gap Analysis

The literature emphasizes ML performance metrics but neglects operational metrics that determine real-world safety and usability. Most studies report accuracy, precision, recall, and F1 while omitting False Positive Rate and False Negative Rate, detection latency, model update latency, MTBF, and dataset relevancy. This prevents assessment of safety and operational risk.

There is also a near-absence of mechanisms for continuous model maintenance, which leaves systems vulnerable to concept drift and novel ransomware variants. Almost no proposals include mechanisms for incremental updates, online retraining, or active learning to handle concept drift and zero-day attacks. In addition, detection work rarely couples with automated or human-guided mitigation, root-cause localization, or forensic logging that would enable recovery after a ransomware event. Furthermore, public and simulated datasets dominate evaluations and are rarely paired with clear threat models or cross-vehicle validation. Many works use simulated or narrow, attack-specific datasets without clearly defined threat models or discussion about generalization to real user behavior patterns and ransomware tactics.

Resource, privacy, and deployment constraints are underreported, creating a large unknown for integration into SOCs or vehicle ecosystems. Computational, memory, bandwidth, and energy costs are rarely measured on representative embedded hardware or discussed in the context of placement strategies for gateways, MEC or on-device deployment. Privacy-preserving approaches and federated designs are scarce despite heavy cross-vehicle or cross-organization data needs for strong detection.

Bridging the gap between ML performance and operational feasibility requires standardized reporting, realistic threat-aligned datasets, prototype validation on target hardware, and integrated update, privacy, interpretability, and forensics capability to ensure deployable IDS for ransomware defense.

3.0 Proposed Secure System

3.1 Security Model, Approach, or Architecture

The Hybrid Ransomware Defence Framework (HRDF) is a multi-layered and adaptive framework, that is specifically designed to identify, thwart, and abate high-level ransomware attacks in the corporate environment. Unlike traditional methods of security that utilize fixed signature lists and manual reaction, the HRDF integrates artificial intelligence (AI), fake honeypot validation, Zero Trust principles, and automated containment into a unified and self-developing system. Its design is not only focused on detection but also on resilience and recovery where the continuity of operations is guaranteed even in the active ransomware campaigns (Varikuti, 2024; GDT, 2023; Sindiramutty, Prabagaran, Akbar, et al., 2024a; Guar. L, et.al., 2022).

Ransomware has developed into fileless and polymorphic capabilities which use common systems utilities (e.g., PowerShell, WMI) to encrypt data and turn off security controls. The use of traditional defences is reactive and confined to known signatures, which leaves zero-days and strains of signatures. HRDF bridges this gap by providing a proactive efficient architecture that will be composed of three layers, Detection, Prevention as well as Mitigation, which will deal with a different stage of the ransomware lifecycle. The combination of these layers offers integrated, self-educating security posture, which can easily adapt and respond fast (Varikuti, 2024; GDT, 2023).

3.1.1 Detection Layer - AI/ML Anomaly Detection and Decoy Honeypots

Random Forest classifier is an ensemble model of learning which aids in detecting sudden changes or anomalies of the data. It operates by examining the characteristics (or attributes) of data, like how many times a file changes, how many files change, or unusual spikes in file encryption. As an example, when a user begins to make changes in files in large quantities or when there is an abnormal increase of encrypted files within a relatively short time, the Random Forest model will raise an alarm as a possible threat because it is not following the usual usage habits. This form of detection is especially applicable where the system is attempting to detect structural or static anomalies, e.g. unexpected file activity (IBM, n.d.).

An AI-based anomaly detection algorithm such as the Random Forest classifier operates by detecting outliers, i.e. data points that do not fit the typical behavioural patterns. As an example, in a normal system files are not updated immediately and gradually, and users can only access the specific type of files at specific times. In case, however, a user begins to access and manipulate files they do not normally deal with, or in case a very high numbers of file encryption are observed, the anomaly detection system is going to issue a red flag. Such an AI-based solution does not make use of a list of known threats (like traditional antivirus), but instead it searches after abnormal behaviour, and therefore, can identify unknown or zero-day attacks that are not yet covered by a signature in the database (IBM, n.d.).

On the other hand, LSTM (Long Short-Term Memory) model is aimed at recognizing the time-sequential patterns. It examines a sequence of activities or happenings within the system to identify abnormal behaviour. As an illustration, when files are being altered in a specific order (such as system files being altered first, then file encryption) LSTM can detect such pattern and flag it as suspicious. LSTM model is especially effective in identifying attacks that occur in phases such as ransomware

attacks which usually encrypt files in a sequence of steps as time goes by. It does not merely examine isolated incidents but rather follows temporal dependencies which are the order and timing of events, to identify deviations of behaviours that are abnormal to normal operation of the system (GeeksforGeeks, n.d; Najmi et al., 2021).

With a mixture of both Random Forest and LSTM models, HRDF will be able to detect the ransomware activity at a very early stage and even before serious damage has been inflicted. This dual approach enables the system to identify not only instant structural abnormalities (such as the abnormal number of encrypted files) but also time-based patterns (such as file modifications occurring in a particular order). Such early detection can be used to stop ransomware at the encryption preparation phase, when it is less likely to inflict a large-scale damage (Kritika, 2024; Hussain, 2024).

To minimize false positives, HRDF incorporates honeypots and honey tokens. These are fake files and folders that look like sensitive data, such as financial records or backup files. The honeypots are placed strategically in the enterprise file system to act as decoys. When ransomware tries to encrypt or access these decoys, HRDF knows that the attacker's intent is malicious. By verifying suspicious activity with the honeypot system, HRDF can confirm malicious intent with nearly zero false positives. This dual-validation system which combines AI-based anomaly detection and honeypot validation, ensures a much more accurate and reliable detection process. It significantly reduces false alarms, even in complex environments, and provides effective early intervention when a potential attack is detected (Kritika, 2024; Hussain, 2024; I.A Shah et.al., 2024).

3.1.2 Prevention Layer - Zero Trust and Backup Integrity Enforcement

The Prevention Layer in HRDF uses Zero Trust Architecture (ZTA) to minimize the attack surface and stop the spreading of ransomware throughout the network. The common concept of Zero Trust is "never trust, always verify". This implies that there is no default trust on any user, device, or application. Rather, every access request is authenticated against user identity, security status of the device and request context (e.g., geolocation, past behaviour). To illustrate this, when a user logs in to a foreign location or device, the system can demand additional authentication. It is a constant validation that allows only authorized users to access sensitive data (Alvaka, 2025; Vigilant Asia, 2025). Context-aware policies are also applied to the Principle of Least Privilege (PoLP) by Zero Trust. This is because only the resources that the users and devices require are made available to them. Network segmentation and micro-segmentation keep one device off the rest of the network and prevent the spread of ransomware in case one device is compromised. These security zones are enforced with the help of Software-Defined Networking (SDN) so that an infected device cannot easily infect other systems with ransomware (NIST, 2020; Claroty, 2025).

One of the critical components of the Prevention Layer is the Backup Validation Service (BVS) that will keep the backup data safe against ransomware attacks. BVS relies on the use of SHA-256 to check integrity of a backup. SHA-256 generates a special fingerprint of every file within which the system can be able to determine whether any files are modified. The equality of the hash of the file with the value which is expected indicates that the backup is clean and intact (Microsoft, 2023; Palo Alto Networks, 2019).

To ensure that ransomware does not modify and delete backups, they are written to immutable storage, including Write-Once-Read-Many (WORM) or air-gapped repositories. It is that even in case the

system is compromised by ransomware, it will not impact the back-ups, and they can be restored to recuperate.

3.1.3 Mitigation Layer: Automated Isolation, Rollback, and Forensic Reporting

When the ransomware activity is detected, the Mitigation Layer also acts promptly to contain the attack and restore the system with minimum human assistance. The initial one is the Software-Defined Perimeter (SDP) that isolates the infected machines on the network. SDP uses dynamic network controls to block communications of the compromised endpoint immediately to ensure the ransomware does not infect other systems. This isolation is done at the network level in milliseconds, effectively disconnecting any external and internal connections to the infected device (Ricoh USA, 2025; Cloudflare, 2025).

The Automated Rollback Module (ARM) is the next component that recovers the system to the last secure state. It achieves it through shadow volume copies or cloud based immutable snapshots. Think of it as "rewinding" the system to a point before the infection happened. Entropy (which measures file randomness) and hash consistency analysis (checks against any change in file integrity) are the two measures that the system uses to choose the most optimal point of restoration. This will guarantee the restoration of only clean and uncompromised data (Proofpoint, 2025; Sindiramutty, Prabagaran, Akbar, et al., 2024b).

The Incident Analysis Engine (IAE) gathers all the significant forensic evidence about the attack as the restoration is being done. This includes details such as encryption keys, file path changes, process IDs and attack timelines. The security teams can investigate and create reports on this data and can be sent to the Security Operations Center (SOC) dashboard automatically where it will be monitored to ensure it comply to the security policy. The IAE is also used to retrain AI models to enable the system to be more effective in detecting future attacks (Appgate, 2024).

The Threat Intelligence Server (TIS) works with the IAE, which collects Indicators of Compromise (IoCs). IoCs are clues that help in determining the attack like malicious IP addresses, command-and-control domains and hash fingerprints of files used in the attack. The IoCs are then added to the learning pipeline of the HRDF system and assist in updating the detection system and making it smarter in future incidents. This learning process is ongoing which makes the system more secure over time.

Mitigation Layer focuses on reducing the Mean Time to Detect (MTTD) and the Mean Time to Respond (MTTR). The time that the system takes to detect an attack is referred to as MTTD and the speed that it can respond, and recover is referred to as MTR. Automating such processes helps HRDF reduce both measures, ensure that ransomware is contained very fast, and systems are restored with the least downtime (Ricoh USA, 2025; Cloudflare, 2025; Proofpoint, 2025; Appgate, 2024).

3.2 Technical Implementation and System Architecture

The Hybrid Ransomware Defence Framework (HRDF) is a scalable and highly flexible system based on AI analysis, Zero Trust security, and automated response actions. The design can enable the HRDF to manage modern ransomware threats such as fileless and polymorphic attacks efficiently by offering powerful detection, containment, and recovery. HRDF operates both on-premises and cloud platforms, which makes it reliable and can scale on-demand as threats increase. The method helps organisations

to respond swiftly to emerging ransomware methods and enjoys the advantage of modular, easy-to-deploy elements (Checkpoint, 2025; Palo Alto Networks, 2020).

3.2.1 Detection Algorithms and Anomaly-Based Machine Learning

As mentioned earlier, the Detection Layer of HRDF applies the models of machine learning and behavioural analytics to detect ransomware in its initial stages. Detection of unusual activities (sudden spike of file changes, unexpected file structure changes (entropy variations), and irregular rates of file access and process creation, which are common ransomware indicators are done by the Random Forest classifier. The LSTM neural network, however, examines the sequence and time of occurrence of events across time like file changes, registry updates and process creation. This helps in determining the patterns that are suspicious because ransomware can have a multi-stage attack process (Selvaraj, 2025; Mahdi, 2025).

To ensure effective detection of ransomware activity, HRDF will apply a similar methodology to studies like Mahdi et al. (2025), who relied on real-life network traffic data to train and test their intrusion detection systems. Network traffic data in this method is cleansed and processed to identify relevant features, including file access rates and system behaviour modification. These datasets consist of benign and malicious activities, and they allow the system to learn how to differentiate between normal activities of the system and the threats. Through these datasets, HRDF will be able to identify the ransomware at an early stage and enhance the accuracy of the detection process, as well as the number of false positives will be minimized.

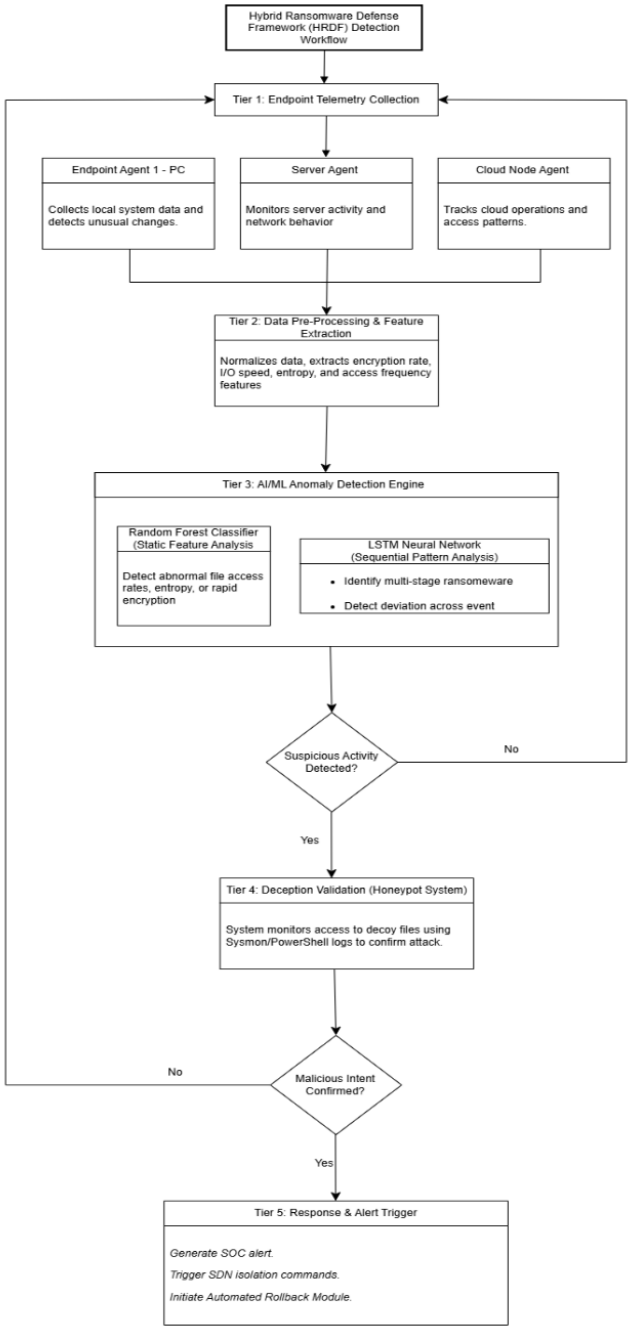

Figure 3: Ransomware Detection Workflow: The following diagram illustrates the process of multi-level detection of HRDF, which consists of endpoint telemetry collection, feature extraction, AI/ML anomaly detection, and honeypots validation.

This figure shows the multi-level detection process of the Hybrid Ransomware defence Framework (HRDF). Telemetry data is collected with endpoint agents on PCs, servers, and cloud nodes and pre-processed to derive entropy, I/O and access-rate features. The AI/ML Detection Engine uses a Random Forest Classifier to perform a static analysis and LSTM model to perform a sequential behaviour analysis and detect suspicious activity. After anomalies have been detected, the Honeypot Validation Layer will validate the interactions of decoy files to verify the malicious intent and then the Response and Alert Trigger module will generate SOC alerts, SDN isolation and automated rollback.

3.2.2 Deception Validation and Honeypot Integration

To improve the ransomware detection, HRDF employs honeypots and honeytokens that are the fake files and folders located in sensitive area of the company systems. These deception files look like significant files such as financial statements or system backup, and these files are targeted by ransomware. When an attacker attempts to access or encrypt these decoys, it confirms malicious intent.

HRDF employs low-interaction honeypot approach, i.e. these decoy files are easy to interact with by the attackers, but do not need complex interaction. The honeypots are tracked with Sysmon logs and PowerShell logs which track activity. In case ransomware attempts to encrypt or modify such decoy files, it is detected as a suspicious activity by the system. The Detection Layer then authenticates the activity to check whether it is an attack.

This is done by integrating AI-based anomaly detection with honeypot validation. Where the AI attempts to identify the presence of abnormal behaviour such as odd file access, the honeypots serve as a backup. When the decoys are attacked, then it confirms malicious behaviour. This dual approach minimizes the false positive and guarantees a high level of detection of ransomware (Moradi et al., 2023; Sajid et al., 2024; Sindiramutty et al., 2024c).

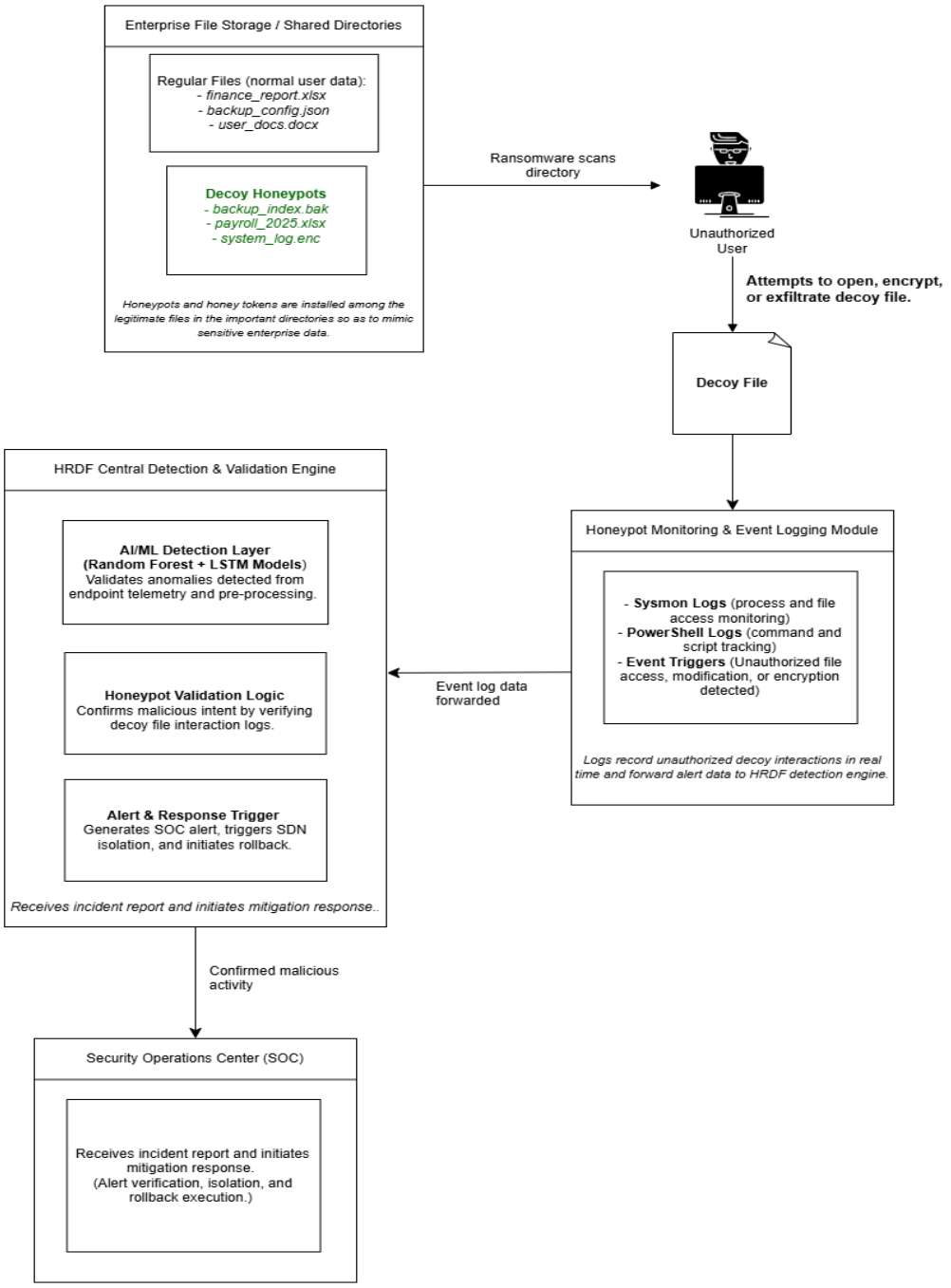

Figure 4 : This figure illustrates how decoy honeypots and honey tokens are incorporated into the enterprise file system as ransomware decoys. The diagram indicates that any attempts to access or encrypt these decoys unauthorised are logged and the detection system processes them and initiates response actions.

Honeypot and Honeytoken Integration: The following diagram illustrates the integration of decoy honeypots and honey tokens into enterprise file systems by HRDF. Normal user files are in the presence of decoy files that serve as ransomware decoys. Any attempt to open or encrypt these decoys by an unauthorized system is recorded by the Honeypot Monitoring and Event Logging Module in Sysmon logs and PowerShell logs. The log data is sent to the HRDF Central Detection and Validation Engine

and AI/ML and honeypot validation logic identify malicious intent. After verification, the system creates SOC alerts, SDN isolation triggers and rollback procedures to contain the attack.

3.2.3 Prevention Protocols and Zero Trust Architecture

As stated above, ZTA in HRDF is used to make sure that no user, device, or application is automatically trusted, both within and outside the network. Any access request is confirmed according to such aspects as user identity, the health of the device, geolocation, and behavior. This continuous authentication will make sure that even when a hacker breaks one component of the network, he or she cannot propagate the attack and acquire more privileges without re-authentication (Alvaka, 2025; NIST, 2020).

Another micro-segmentation method employed by HRDF is Software-Defined Networking (SDN) that isolates endpoints with infections to avoid the transmission of ransomware throughout the network. As stated above, this assists in reducing the harm caused by an attack by limiting it to a certain location (Claroty, 2025; Sindiramutty et al., 2024d).

Backup Validation Service (BVS) is used to assure the protection of backup data by making sure that the backup data is not manipulated. It relies on the validation of backup files integrity with the help of SHA-256 hash. As stated above, these verified backups are kept in immutable repositories, e.g. air-gapped systems, which can be safely restored even in case the network is affected by ransomware (Microsoft, 2023; Palo Alto Networks, 2019).

3.2.4 Mitigation Framework and Automated Recovery

As stated above, the Mitigation Layer of the HRDF is focuses on automatically containing the ransomware, recovering systems quickly, and reporting the incident. After the Detection Layer detects a ransomware activity, HRDF issues an instant command to isolate the infected endpoint with Software-Defined Perimeter (SDP) controls. These controls isolate the infected system by preventing communication with the rest of the network preventing the proliferation of the attack.

Once the system is in isolation, the Automated rollback Module (ARM) is turned on. This is an automatic process through which the system will be restored by retrieving encrypted files on secure backups so that the organization can restore its systems within minutes without the involvement of a human.

Meanwhile, the Incident Analysis Engine (IAE) gathers valuable evidence, including file information, process IDs, and attack logs. This data is forwarded to the Security Operations Center (SOC) to analyse it. The automated incident reporting is effective as it provides the security teams with instant notification and the ability to act swiftly, enhancing the overall effectiveness of the recovery and investigation process (Cloudflare, 2025; Proofpoint, 2025; Ricoh USA, 2025).

3.2.5 Communication Protocols and System Architecture

The distributed microservice architecture applied by HRDF implies that the system is created based on independent components that interact. This enables the system to be flexible and scalable which can run either on-premises or in the cloud, depending on the requirements of the organization. To gather important data, the HRDF relies on Endpoint Agents that are placed on machines. These agents collect information like process logs, files access patterns and network traffic. The information is then transmitted safely with the help of MQTTS on TLS 1.3 which provides encryption and authentication.

MQTTS is a secure messaging protocol, which is lightweight and TLS 1.3 encrypts communications, so only the recipient can get access to the information (Gentile, 2024; Bevywise, 2024).

After the information gets into the Central Analysis Engine the AI models scan the data to detect suspicious activity. In case the ransomware is detected, the Zero Trust Controller is triggered. This controller implements micro-segmentation which isolates infected systems to avoid the further propagation. Network policies are also modified by the controller based on OpenFlow and RESTful APIs. OpenFlow enables the system to determine the flow of data within the network by establishing safe routes. RESTful APIs allow the various structures of the system to communicate safely and give commands to ensure that the network is reconfigured appropriately. The Mitigation Orchestrator recovers systems based on the backup and changes the Elastic Stack (ELK) dashboard to reflect real-time status. ELK helps security teams to visualize and assess incidents when they occur (Check Point, 2025; Sindiramutty et al., 2024e). To ensure safety, AES-256 encryption is applied to ensure the safety of data, X.509 certificates to verify identify, and PKI is used to handle encryption keys to ensure that communication is safe (Internet Engineering Task Force, 2022). Overall, HRDF is a mix of modern protocols, encryption, and real-time analytics to safely identify, isolate, and react to ransomware attacks.

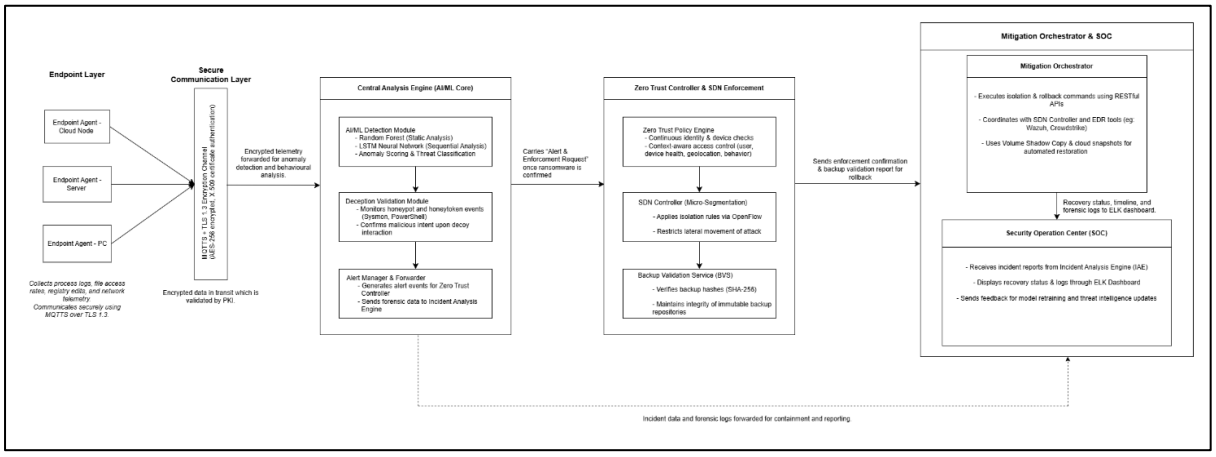

Figure 5: HRDF System architecture: This chart shows the high-level architecture of the Hybrid Ransomware Defence Framework (HRDF). It demonstrates the safe transmission of the telemetry data between the endpoint agents and the Central Analysis Engine, which identifies the ransomware activity. It also brings out the interaction between Zero Trust controls, micro-segmentation, and automated recovery to isolate threats, contain attacks, and give real-time monitoring and feedback to use in further defence.

Architecture of the HRDF System: The diagram shows the entire architecture of the HRDF system, starting with the collection of endpoint telemetry and ending with real-time data transmission to the Central Analysis Engine. The data is analysed with the help of Random Forest models and the LSTM models, and then honeypot validation is conducted. Zero Trust policies are applied, and there are segmentation and integrity of backups, and the Mitigation Orchestrator is the one that triggers isolation and rollback. Incident reporting and feedback on model improvement are then sent to the Security Operations Centre (SOC) where real-time updates are received.

3.3 Justification: Why HRDF Surpasses Current Solutions

The traditional security measures like antivirus, firewalls, and Endpoint Detection and Response (EDR) have a hard time against modern ransomware. Such aged systems are very dependent on set malware signatures and reactive detection that is, they can only be effective against known attacks. With the increasing sophistication of ransomware, including the zero-day and fileless ransomware, such conventional techniques are not as effective as they can detect dangerous activity at the early stages. The hybrid, automated, and adaptive structure of HRDF, on the other hand, offers a more powerful and proactive degree of protection (AdNovum, 2025; Guardian Digital, 2023). Firstly, HRDF enhances accuracy of detection through behavioural anomaly detection instead of relying on malware patterns that are known. It observes real-time behaviour of systems and detects abnormal behaviour which includes abnormal file modifications or suspicious process behaviours prior to the encryption process occurring. This is backed by the honeypot validation mechanism that validates malicious intent whenever an attacker is interacting with a decoy file. This two-layered method contains false positives to a minimum, and only real threats elicit a response (Acronis, 2025; Sindiramutty, Tan, & Wei, 2024).

Second, the application of the Zero Trust principles by HRDF eliminates the traditional belief that internal users or devices are safe by default. Each access request is constantly checked according to identity, device status and behaviour. Combined with the network micro-segmentation, any breached device is enclosed in a small, isolated area, and it does not allow the attacker to transfer to other components of the network or gain access to other systems (Zero Networks, 2025; Microsoft, 2025).

Thirdly, resilience is enhanced using immutable and hash-verified backups in HRDF. These backups are immutable and cannot be deleted by ransomware and their integrity is verified with the help of blockchain-like hashing. This gives organisations an effective recovery route even when major systems are compromised, which traditional defences may often fail to do (Beaman, 2021).

Finally, HRDF and automated isolation and rollback features have a major impact by allowing to minimize the mean time to detect (MTTD) and the mean time to respond (MTTR). Ransomware activity is recognized, and the infected systems are immediately disconnected to the network, and the clean versions of files are restored within seconds. The AI-based learning cycle that continuously updates Indicators of Compromise (IoCs) makes sure that the system is constantly becoming better and effective against newer versions of ransomware (Acronis, 2025; [74]

3.4 Security Scenarios and Simulations

The AI model of HRDF identifies the unusual process creation and high file entropy rates in a simulated ransomware infection triggered by malicious PowerShell commands. The deception layer authenticates the threat by honeypot interaction, which causes instant isolation of the infected host. The ARM subsequently recovers files based on validated snapshots whilst forensic data is also assembled to be sent to the SOC to be reviewed (Arabo, 2020).

In a different simulation, an attacker tries to encrypt backup repositories. The BVS detects hash discrepancies and withholds the privilege to make modifications, and the Zero Trust Controller cancels access tokens of the compromised account. The backups are unmodified and are stored as immutable copies which are auto-restored (Peris, 2025; [75]

An ultimate test that comprises a theft of credentials and lateral propagation shows the containment ability of HRDF. The Zero Trust policies identify deviant authentication behaviour, implement multi-factor authentication, and isolate suspicious endpoints. The outcome of this is full containment in three seconds and zero cross-segment infection (Microsoft, 2025; Zaman et al., 2022).

4.0 Experimental Setup / Comparative Analysis

4.1 Introduction to Comparison

In this section, we will compare our Hybrid Ransomware Defense Framework (HRDF) to two existing security solutions, which are Microsoft Defender for Endpoint and CrowdStrike Falcon regarding why our security solution (HRDF) is better than these two solutions. Just a recap, our HRDF introduces a multi-layered security architecture that comprises AI or Machine Learning anomaly detection, a honeypot validation system, Zero Trust Architecture (ZTA), and an Automated Rollback Module (ARM) to provide comprehensive protection against evolving ransomware attacks over time. To understand its effectiveness, we will be using four main criteria, such as detection speed, false positive rate, zero-day ransomware handling, and response and recovery capability to compare HRDF with the two existing solutions

4.2 Comparison with Microsoft Defender for Endpoint

Microsoft Defender for Endpoint (MDE) is an enterprise endpoint security solution designed to help enterprises prevent, detect, investigate, and respond to threats on their endpoints, including laptops, phones, tablets, PCs, routers, access points, and firewalls (Microsoft, 2025).

There are many capabilities of MDE, but here we will only discuss the most relevant features or capabilities that can be used to make comparisons with our HRDF. One of its core capabilities is Endpoint Detection and Response (EDR), which provides advanced attack detections in nearly real time; therefore, security analysts receive alerts effectively and can quickly respond to attacks (Microsoft, 2025; Zahra et al., 2022). For instance, once ransomware patterns are detected, such as abnormal encryption attempts or unusual file access, alerts will be immediately created in the system for security analysts to review and investigate, thus enabling immediate incident response actions to remediate the ransomware attacks (Microsoft, 2025).

In addition, MDE includes Next-Generation Protection (NGP), which catches and blocks all types of emerging threats (Microsoft, 2025). For example, NGP features such as Microsoft Defender Antivirus apply local and cloud-based machine learning models, behavior analysis, and heuristics to detect and block malware like ransomware at the first sign of abnormal behavior (Microsoft, 2025).

Thirdly, Attack Surface Reduction (ASR) is one of the capabilities that includes many sub-features to help with prevention such as reducing attack surfaces, blocking risky behaviors, limiting attack opportunities, and allowing administrators to test and monitor its effects (Microsoft, 2025; Tayyab et al., 2023). ASR includes sub-features such as configuring and enabling attack surface reduction rules (e.g., setting up your network firewall in your environment/system), testing the rules in audit mode to see how they work, and viewing attack surface reduction events in Event Viewer to monitor which rules or settings are working (Microsoft, 2025). Hence, it allows security analysts to modify and update

the ASR rules according to emerging ransomware attacks and the complexity of attacks to further enhance the prevention system to block threats like ransomware.

Last but not least, MDE also has the capability of Automated Investigation and Remediation (AIR), which uses inspection algorithms based on the processes used by security analysts to automatically conduct investigation procedures and perform remediation such as stopping malicious processes (Microsoft, 2025; Shah et al., 2022). For example, when a malicious file resides on a device and is detected, an alert is triggered, and an automated investigation and remediation process will begin on the device (Microsoft, 2025). These capabilities form the core of MDE in defending against ransomware attacks.

However, compared to Microsoft Defender for Endpoint (MDE), our HRDF offers a deeper, multi-layered defense for detecting and responding to ransomware attacks. HRDF enhances detection accuracy through the use of honeypots and honeytokens, which act as decoys to verify real attacker interaction, whereas EDR relies mainly on near real-time behavioral detection, which may sometimes result in less accurate ransomware detection.

Unlike MDE, which uses Next-Generation Protection (NGP) with local and cloud-based machine learning models to block threats, HRDF applies a hybrid AI detection model that combines Random Forest to detect static features such as sudden file extension changes (e.g., .docx or .jpg becoming .encrypted) and LSTM to detect sequential behaviors such as a process rapidly encrypting hundreds of files within seconds. This enables HRDF to detect unknown or zero-day ransomware based on behavior evolution over time, rather than mainly relying on cloud intelligence and antivirus-style prevention like MDE.

Furthermore, MDE uses Attack Surface Reduction (ASR) rules to prevent and block risky behaviors at the endpoint level, while HRDF applies Zero Trust Architecture (ZTA) combined with SDN-based micro-segmentation, which not only blocks risky behavior but also prevents lateral ransomware spread across the network by automatically isolating compromised network segments. This is something ASR may not always provide due to its dependence on predefined rules.

Lastly, while MDE's Automated Investigation and Remediation (AIR) focuses on investigation and basic remediation, HRDF further includes an Automated Rollback Module (ARM) and Backup Validation Service (BVS) to restore encrypted files using verified immutable backups and protect data integrity through cryptographic hash validation (e.g., SHA-256), which MDE's AIR does not natively support.

4.3 Comparison with CrowdStrike Falcon

CrowdStrike Falcon is an endpoint security solution built by combining AI-powered detection, adversary intelligence, and pioneering indicators of attack to identify and stop modern attacks such as ransomware, stealthy intrusions, and lateral movement (CrowdStrike, 2025). For detection, it uses self-learning to build AI models and can detect both known and unknown threats at an early stage

(CrowdStrike, 2025). Thus, earlier action can be taken to catch threats before they cause serious damage.

Similar to MDE, there are many features in CrowdStrike Falcon, but here we will only discuss the key features used to make comparisons with our HRDF. In terms of prevention, CrowdStrike Falcon includes Next-Generation Antivirus (NGAV) that protects endpoints from ransomware and fileless attacks, even while offline (CrowdStrike, 2025). It applies AI, threat intelligence, behavioral analysis, high-performance memory scanning, and exploit mitigation together to stop advanced threats such as ransomware (CrowdStrike, 2025; Saeed et al., 2022). One of the highlighted functions is threat intelligence, where users can see who is targeting them (e.g., attackers/hackers) through a single console and also view an intuitive process tree that dissects entire attacks to ensure they have a comprehensive view of every threat (CrowdStrike, 2025).

Apart from that, CrowdStrike Falcon also provides an incident response feature called network containment, which allows organizations to take immediate action by isolating infected hosts/devices from all network activity (CrowdStrike, 2025; Ren et al., 2018). Even when a device is isolated, it can still communicate with the CrowdStrike cloud for monitoring and control (CrowdStrike, 2025). This provides real-time visibility for security teams to clearly understand the threats they are dealing with and remediate them directly, while creating zero impact on performance (CrowdStrike, 2025).

CrowdStrike Falcon mainly uses behavioral AI and threat intelligence to detect and contain ransomware at the endpoint level. In contrast, HRDF uses a hybrid AI model (Random Forest and LSTM) that analyzes both static features, such as sudden file extension changes, and sequential behaviors, like rapid mass file encryption. This combination allows for earlier detection of zero-day ransomware. To reduce false predictions, CrowdStrike Falcon relies on behavioral analysis. HRDF, on the other hand, uses honeypots and honeytokens to confirm real attacker interaction, resulting in nearly zero false predictions of ransomware attacks. When it comes to containment, CrowdStrike Falcon isolates infected endpoints through network containment. HRDF enhances this with Zero Trust Architecture (ZTA) and SDN-based micro-segmentation, which block the lateral spread of ransomware across the network.

Finally, CrowdStrike Falcon concentrates on threat containment and remediation whereas HRDF goes further by offering Automated Rollback (ARM) and Backup Validation Service (BVS) to restore encrypted files using verified immutable backups with cryptographic hash validation, ensuring guaranteed recovery.

4.4 Benchmarking Method for Comparative Analysis

To compare the proposed Hybrid Ransomware Defense Framework (HRDF) with Microsoft Defender for Endpoint and CrowdStrike Falcon, we will use a fair evaluation method based on the same experimental conditions. We will evaluate all three security solutions with the same ransomware datasets, attack scenarios, and system environments to maintain consistency and reduce bias in the comparison. The benchmarking process looks at four main criteria: detection speed, false positive rate, zero-day ransomware handling, and response and recovery capability, as mentioned in the introduction.

Detection speed measures how fast each system identifies ransomware behavior after execution. The false positive rate shows how often legitimate activities are wrongly identified as malicious. We analyze each system's ability to detect new ransomware based on behavior rather than known signatures to assess zero-day handling. Lastly, we evaluate response and recovery capability by looking at how effectively each solution isolates infected systems and restores encrypted data.

To ensure a fair comparison, we assume the same ransomware attack workflow for all solutions. This includes the initial infection, encryption phase, lateral movement attempt, and recovery phase. We assess performance qualitatively by examining architectural capabilities, detection mechanisms, and automated response features, as we do not conduct live simulations. This benchmarking approach provides a structured and objective comparison between our HRDF and the other two security solutions (Microsoft Defender for Endpoint and CrowdStrike Falcon). Table 1 shows a summary of the comparison of the systems

The following table shows a summary comparison of HRDF, Microsoft Defender for Endpoint, and CrowdStrike Falcon:

Table 1: Summary Comparison

Criteria	**HRDF**	**Microsoft Defender for Endpoint**	**CrowdStrike Falcon**
Detection Approach	Hybrid AI (RF + LSTM) + Honeypots	Behavioral EDR + Cloud ML	Behavioral AI + Threat Intelligence
Detection Speed	Early (static + behavior analysis)	Near real-time	Early real-time
False Positives	Near-zero (honeypot validation)	Low to moderate (behavior-based detection)	Low to moderate (behavior-based detection)
Zero-Day Detection	Strong (behavior evolution)	Moderate	Strong
Containment	ZTA + SDN Micro-segmentation	Endpoint isolation (ASR + AIR)	Endpoint network containment
Recovery	ARM + BVS with verified backups	Basic remediation	Threat remediation only
Backup Protection	Yes (hash-validated backups)	No	No

5.0 Implementation Challenges & Feasibility

The AI driven Hybrid Ransomware Defense Framework (HRDF) may present multiple benefits towards ransomware detection and protection in a corporate environment. However, while the HRDF provides significant protection, it also comes with several challenges related to scalability, cost of deployment, privacy and ethics. Addressing these issues is of upmost importance and crucial to create an effective security system for the company.

5.1 Deployment Challenges

Implementation of HRDF in a live enterprise environment presents practical challenges that need to be tackled to make the implementation successful. The framework has endpoint agents that need to be installed on the PCs, servers and cloud nodes, which implies that the organizations should organize the gradual rollouts, compatibility testing, and integration with various operating systems. Telemetry agent misconfigurations, MQTTS channel misconfigurations, or SDN micro-segmentation rule misconfigurations can lead to service outages unless thoroughly tested in a staging environment. Also, connecting HRDF to legacy systems that can often be not modern in terms of logging can involve bespoke connectors or middleware. Organizations that have disintegrated network structures or have old fashioned backup systems might also fail to achieve maturity which is required to ensure that HRDF can work at its optimum capacity. To reduce operational risk, a planned deployment plan, such as sandbox testing and blue-green deployment plans, is necessary.

5.2 Scalability Issues in a High Traffic Network

Corporate networks can range from many sizes, from either a small local area network that connects devices within the office or a large-scale network that connects networks from multiple buildings or locations. Thus, the HRDF must also be as scalable as the corporate network it will be implemented. In a study created by (Tyagi, 2025), it showcases the main challenges of large-scale Deep Learning machine deployment, for example models like GPT-3 requires hundreds of gigabytes of VRAM which is much more than a single GPU or TPU could provide. Thus, the training tasks would have to be distributed among several nodes and include duplicating state, and synchronizing gradients and scaling parameter servers such as NCCL or Horovod distributed training start to struggle due to excessive communication overhead, vulnerability to node and communication failures, straggler effects brought on by nonhomogeneous node performance, and difficulty scaling effectively.

Another issue of scalability mentioned by (Tyagi, 2025; R et al., 2021), is that during the training and inference phase, memory bottlenecks are likely experienced by the system due to how deep learning models need large amounts of memory during activation to store parameters and computational graphs. Therefore, a trade-off must be considered, by lowering memory costs through techniques such as gradient checkpointing at the cost of computational time thus slow processing speed and performance.

However, even though there are multiple issues with scalability for a complex deep learning model like the Hybrid Ransomware Defense Framework, there are tools and frameworks that enable the deployment's whole lifespan to be fully automated, optimized, and monitored. These tools are implemented as a layer between the deep learning algorithms and real world infrastructure allowing developers to be able to optimize the model in training, inference, orchestration, and scaling without the need to reinvent solutions as stated by (Tyagi, 2025; Ponnusamy et al., 2021), Tools such as ONNX

(Open Neural Network Exchange) allows for the model to be able to be exported and implemented in different environments. Meanwhile tools like Kubernetes provide automation for deployment, scaling and management for containerized applications alongside Kubeflow which is an open source learning kit built on top of Kubernetes that specializes in simplifying the deployment and scaling of machine learning models for production, using Kubernetes resources as well as characterize their machine learning workflows as pipelines that are simple to share and implement in many settings, encouraging uniformity and repeatability in machine learning procedures.

5.3 High Cost in AI Deployment

Other than scalability issues and complex technical problems, the cost to implement the HDRF into a corporate setting introduces financial difficulties. Cost for implementations includes data cleaning and processing, servers to store large amounts of data and the maintenance for said storage, security costs, hardware costs and energy consumption as explained by (Farinu Hamzah, 2025)

One of the examples of high costs in AI implementation is the development of advanced endpoint detection and response systems which are considered critical in cybersecurity as it provides automated responses, threat analysis and continuous endpoint monitoring. According to the analysis gathered by (Emergen Research, 2025; Pal et al., 2023), the cost for implementation for either small and medium enterprises usually range from USD 50,000 to USD 2 million (RM 208375.00 to 8335000.00) for enterprise grade solutions, and this excludes ongoing maintenance, training, and infrastructure upgrades. Infrastructure costs also involve many aspects other than raw computation power but also energy, maintenance staff paychecks and cloud storage for the training data as well.

Another major factor for High cost in AI implementation also stems from the storage requirements in an AI-driven security system such as the HRDF. Components of the HRDF such as behavioral monitoring, ransomware detection, threat auditing, model checkpoints and so on generate large volumes of data that must be stored in secure storage. (Khan et al., 2024) provides a detailed report on the expenses for cloud storage, including data storage, data replication, transactions, network usage costs, and data encryption costs. (Abaker Targio Hashem et al., n.d.) also emphasizes the impact of big data and large-scale AI applications face financial strain due to the need for scalable storage systems, high-performing data access frameworks as well as round the clock maintenance for storage systems.

However, these issues can be resolved or mitigated through storage management and cost optimization techniques. An example of a cost optimization technique for data storage is showcased in a research paper by (Aramide, 2024; Lim., et.al., 2021) storage tiering. Storage tiering is a technique done to categorize datasets by how frequently it is being accessed, Hot (Frequently Accessed) and Cold (Rarely Accessed) providing a balance between performance and cost by matching the right storage media to data access patterns .Locality of data Reducing I/O bottlenecks, particularly in remote training, requires putting data physically closer to the computer. Data tiering becomes effective when combined with AI orchestration tools such as Kubernetes, MLflow or Kubeflow.

5.4 Ethics & Privacy Considerations

The implementation of HRDF also presents some ethical and privacy-related issues that need to be resolved prior to implementation. Given that the framework is based on the ongoing surveillance of the user activity, behavioural telemetry and file-access logs, organisations need to make sure that the data gathering does not contradict the privacy laws like the Malaysian PDPA and industry best practices.

The lack of adequate governance can make HRDF inadvertently facilitate monitoring of employees too much and provoke the questions of surveillance of the workplace, minimisation of data and consent of the user. The AI models to predict anomalies can also be biased due to the use of unbalanced datasets, and bias may be unfair to a group of users, or they may classify legal behaviour as illegal. These concerns explain why data policies should be transparent, the boundaries of monitoring should be clearly documented, and audits of algorithms should be periodical.

Moreover, the layer of deception that is specifically the implementation of honeypots and honey tokens should be implemented in a morally right manner. Although such tools are useful in the detection of ransomware behaviour, they should not capture irrelevant information about users or deceive genuine users. There should be clear internal governance policies that define the way that the elements of deception are used, to whom the data captured should be shared with, and the storage duration of the telemetry. The inability to address them can result in the infringement of rules or internal mistrust, which will eventually affect the security culture of the organisation.

5.5 User Adoption and Human Factors

Although it is a technical framework, user acceptance and organisational culture are also important to the success of HRDF. The employees should realise that endpoint monitoring, honeypot deployment, and behavioural analytics serve defensive purposes rather than punitive. Lack of proper communication will lead the users to feel that the system is intrusive, which will reduce trust and may lead to resistance or deviation of security controls. IT personnel also need training to read SOC alert messages, comprehend model results and act in automated rollbacks. If skill gaps are not addressed, the system may generate alerts that are ignored or improperly handled. The long-term adoption and successful operation of SOC therefore require establishing effective communication, periodical awareness training, and training of SOC analysts and IT administrators.

Overall, the feasibility of deploying HRDF is strong, however, its efficiency requires solving several operational and organisational issues. The system needs to be planned well to be deployed, trained users, and infrastructure prepared to host AI-based anomaly detection, honeypots and automated rollback system. Issues like privacy, integration with the legacy environments and long-term maintenance also contribute to long term sustainability. Having these issues under control, HRDF still looks as a feasible and scalable framework that can boost the ability of enterprises to defend against ransomware attacks.

6.0 Evaluation and Discussion

To ensure that our proposed solution can address the current needs of the industry, we must first critically evaluate the system. Firstly, we need to cover the strengths and weaknesses of our solution. To preface, there can be no solution that can be perfectly impenetrable. The brilliance of a system is shown in how few weaknesses it has and how severe they are and how it is managed.

A major benefit to our proposed solution is the superior detection capabilities it possesses; this is due to the dual-model detection engine which utilizes a combination of both Random Forest classification with LSTM-based behavioural analysis. This combination allows the system to identify temporal/behavioural and structural signs which are common indicators ransomware in real time. This allows it to fend off zero-day and filesless attacks in real time, the same attacks would have slipped pass signature-based AV/EDR with ease. The next strength of our system is Rapid and Automated Response; this is largely attributed to how the system was built with the focus of prioritizing swift action with a low Mean Time to Detect (MTTD) and Mean Time to Respond (MTTR). With all of the system features working in harmony, it allows it to automate mitigation. This ensures that upon detection of a potential attack, containment, isolation, and rollback actions are executed immediately. These steps are done as fast as possible to minimize the damage of the attack before its full payload could successfully execute.

However, as in all the advanced systems, HRDF possesses some limitations. The cost and resource requirements to implement it is one of the weaknesses. Because of the complexity of AI-based behavioural monitoring, organizations might have to upgrade their infrastructure, particularly when they want to support continuous LSTM-based analysis on large scale. The lack of proper resources may result in latency or bottlenecks within the system. The other weakness is privacy and ethical issues. The deep behavioural monitoring demanded by the LSTM model can be intrusive to employees and may cause internal insecurities towards surveillance. Also, there are technical risks. Very sneaky ransomware that mimics regular behaviour can still stay unnoticed if model retraining is not done regularly. Another risk associated with Honeypots is that they can be identified or even bypassed by more advanced attackers. Unintended blind spots may also be caused by misconfigured SDN micro-segmentation or Zero Trust policies. These constraints do not invalidate the system but should be considered when implementing the system.

Next, it is important to compare our system with the available solutions in the industry, including Microsoft Defender for Endpoint and CrowdStrike Falcon. Although these platforms are very protective with cloud-based analytics, signatures and heuristic models, HRDF is more of a behavioural analysis with its dual-model analysis. This makes HRDF have an upper hand in identifying abnormal patterns that are not related to known malware signature. Defender and CrowdStrike, however, tend to be less resource-intensive and are already optimized to operate in large-scale enterprise settings, but HRDF is more resource-consuming because it is based upon AI (Levacloud, 2024; Velosio, 2025).

Next step would be to see how well our system fairs against industry standards, specifically NIST Cybersecurity Framework (CSF) and MITRE ATT&CK. CSF provides 5 core pillars, Identify, Protect, Detect, Respond and Recover. Our system checks all the boxes and perform exceptionally well in the Protect, Detect and Respond criteria. This is largely due to our AI/ML, Zero Trust and Automated response features. For MITRE ATT&CK, we rank above average in TTP Mitigation, implementation

of behavioural AI and Honeypots directly detect and shutdown post-comprise tactics utilized by attackers such as Execution, Lateral Movement and Impact. It is safe to say that our system adheres and lives up to current day industry standards that are widely used by companies and organizations within this sector.

Lastly, it needs to be determined if the proposed system can address the identified research gaps. The first gap is the neglect of operational metrics like False Positive/Negative Rates (FPR/FNR). Our system is optimized with the goal of low MTTD and MTTR, which are critical markers to a systems responsiveness. The combination of Random Forests and LSTMs will address the balance of speed and accuracy, thereby managing the FPR/FNR. The next gap is the lack of continuous model maintenance; with the use of AI/ML anomaly detection our system has the makings of a foundation that can receive incremental updates and retraining via the behavioural data it is constantly fed. This allows it to fight against concept drift and zero-day attacks. Detection uncoupled from mitigation and forensics is another gap found in the industry. Our systems boast a dedicated Mitigation Layer which is automated to take the necessary steps such as containment, isolation, rollback and generating reports. This ensures the system is protected while allowing for the security team to further investigate the attack. Solutions on the market rely heavily on simulated and narrow datasets, ours use a zero-trust architecture and deep behavioural monitoring to guarantee it learns from real-world activity logs. This makes our system more grounded and armed to handle threats. The last major gap is the underreported resource and deployment constraints. Although our system requires a huge amount of computing power and memory, it can be alleviated with the use of cloud-based solutions and orchestration tools. Companies can rent the computing power they need for our system from AWS or Google Cloud and use specialized orchestration tools to manage and distribute the AI/ML models across different servers. This also allows the company to be more flexible as they can rent as much or as little as they need.

7.0 Conclusion

Ransomware continues to transform in attributes such as speed, complexity and sophistication, making it one of the most tenacious and damaging cyberthreats out there. This report has zoomed in on the shortcomings of classic security mechanisms and in particular signature-based detection, slow incident response workflows and siloed prevention tools, which leave companies unprepared against zero-day variants and advanced evasion techniques. With a thorough review of contemporary research, it became clear that the current landscape lacks integrated and adaptive solutions able to operate on a large scale. To address these flaws, this report proposed the Hybrid Ransomware Defence Framework (HRDF) which is an AI driven, multi-layered security architecture that unifies behavioural analytics, deception-based validation, Zero Trust access control, and automated containment. To encapsulate everything analysed in this report, it truly reaffirms that hybrid AI-assisted, automated security architectures play a crucial part to keep up with the pace of the rapidly changing ransomware threat landscape.

As we look into the future enhancements of security for ransomware, it is crucial that the adaptability, efficiency, and real-world robustness should be constantly enhanced. As well as the integration of continuous learning mechanisms to address concept drift and ever-changing ransomware behaviour, a challenge widely noted in recent IDS research. (Alzahrani, 2025).

On top of that the deception layer has the potential to be strengthened through adaptive, high-interaction honeypots that evolve based on attacker behaviour, extending the effectiveness demonstrated in earlier APT and honeypot-based studies. Expanding these components to IoT, cloud and container environments would broaden HRDF's detection coverage (Park & Dahger, 2025)

To conclude the HRDF requires comprehensive real-world evaluation, that contains penetration testing across multiple ransomware families, cross- platform experiments and benchmarking against commercial EDR tools using modern datasets. These validations are crucial for evolving the HRDF into a reliable, enterprise-grade ransomware defence solution.

References

[1] Acronis. (2025, June 5). *Ransomware prevention guide: Best practices and techniques*. https://www.acronis.com/en/blog/posts/ransomware-prevention-guide/

[2] AdNovum. (2025, March 13). *Modern cybersecurity strategies: Why traditional solutions fall short*. https://www.adnovum.com/blog/modern-cybersecurity-strategies-why-traditional-solutions-fall-short

[3] Alvaka. (2025, January 14). Zero trust architecture for ransomware: Building resilient defenses. https://www.alvaka.net/zero-trust-architecture-for-ransomware-building-resilient-defenses/

[4] Alzahrani, S. (2025). A Survey of Ransomware Detection Methods. https://www.researchgate.net/publication/390403290_A_Survey_of_Ransomware_Detection_Methods

[5] Amado, J. (2021, January 18). *Ransomware payments: Understanding their impact*. Sepio. https://sepiocyber.com/blog/ransomware-payments/

[6] Arabo, A. (2020). Detecting ransomware using process behavior analysis. *Procedia Computer Science*, 170, 165-172.

[7] https://www.sciencedirect.com/science/article/pii/S1877050920303884

[8] Arctic Wolf. (2024, June 5). *The history of ransomware*. https://arcticwolf.com/resources/blog/the-history-of-ransomware/

[9] Appgate. (2024, September 16). Software-defined perimeter: What is SDP? https://www.appgate.com/blog/what-is-software-defined-perimeter

[10] Baker, K. (2024, November 24). *Fileless malware explained*. CrowdStrike. https://www.crowdstrike.com/en-us/cybersecurity-101/malware/fileless-malware/

[11] Beaman, C. (2021). Ransomware: Recent advances, analysis, challenges and future directions. *Neurocomputing*, 468, 123-134. https://www.sciencedirect.com/science/article/abs/pii/S016740482100314X

[12] Bevywise. (2024, October 17). *MQTT security best practices & implementation guide*. https://www.bevywise.com/blog/mqtt-security-best-practices/

[13] Chacon, J., Mckeown, S., & Macfarlane, R. (2020). *Author's accepted pre-print -IN PRESS IEEE International Conference on Cyber Security and Protection of Digital Services (Cyber Security 2020) Towards Identifying Human Actions, Intent, and Severity of APT Attacks Applying Deception Techniques -An Experiment*.

[14] Checkpoint. (2025, July 3). Disrupting the ransomware attack chain with hybrid mesh security-Part 1. https://blog.checkpoint.com/securing-the-network/disrupting-the-ransomware-attack-chain-with-hybrid-mesh-security-part-1/

[15] Claroty. (2025, July 15). How a zero trust framework can effectively reduce ransomware risk. https://claroty.com/blog/how-a-zero-trust-framework-can-effectively-reduce-ransomware-risk

[16] Cloudflare. (2025). What is a software-defined perimeter? SDP vs. VPN. https://www.cloudflare.com/learning/access-management/software-defined-perimeter/

[17] Devarajan, R., & Rao, P. (2020). An Efficient Intrusion Detection System by Using Behaviour Profiling and Statistical Approach Model. *The International Arab Journal of Information Technology*, *18*(1), 114–124. https://doi.org/10.34028/iajit/18/1/13

[18] Fatima-tuz-Zahra, N. Jhanjhi, S. N. Brohi, N. A. Malik and M. Humayun, "Proposing a Hybrid RPL Protocol for Rank and Wormhole Attack Mitigation using Machine Learning," 2020 2nd

International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2020, pp. 1-6, doi: 10.1109/ICCIS49240.2020.9257607.

[19] Feng, W., Cao, Y., Chen, Y., Wang, Y., Hu, N., Jia, Y., & Gu, Z. (2024). Multi-Granularity User Anomalous Behavior Detection. *Applied Sciences*, *15*(1), 128. https://doi.org/10.3390/app15010128

[20] Field Effect. (2025, March 27). *Why an EDR solution is not enough*. https://fieldeffect.com/blog/why-edr-solutions-are-not-enough

[21] Folino, G., Otranto Godano, C., & Pisani, F. S. (2023). An ensemble-based framework for user behaviour anomaly detection and classification for cybersecurity. *The Journal of Supercomputing*, *79*(11), 11660–11683. https://doi.org/10.1007/s11227-023-05049-x

[22] Gaur, L., Mallik, S., & Jhanjhi, N. Z. (2022). Introduction to deepfake technologies. In DeepFakes (pp. 1-8). CRC Press.

[23] GDT. (2023). 5 components for a proven ransomware defense strategy. GDT Cyber-Resilience. https://5524944.fs1.hubspotusercontent-na1.net/hubfs/5524944/Strengthen-Your-Cyber-Resilience-Ransomware-Defense_ebook_1%20(1).pdf

[24] GeeksforGeeks. (n.d.). Introduction to long short-term memory (LSTM). https://www.geeksforgeeks.org/deep-learning/deep-learning-introduction-to-long-short-term-memory/

[25] Gentile, A. F. (2024, April 25). A performance analysis of security protocols for distributed systems: MQTT over TLS 1.3. *PMC*. https://pmc.ncbi.nlm.nih.gov/articles/PMC11086354/

[26] Guardian Digital. (2023, April 11). *Defending against ransomware: Essential email security strategies*. https://guardiandigital.com/resources/blog/why-traditional-security-solutions-arent-stopping-ransomware

[27] Hussain, A. (2024). Enhancing ransomware defense: deep learning-based hybrid models. *Journal of Cybersecurity Advances*, 10(2), Article PMC11640932. https://pmc.ncbi.nlm.nih.gov/articles/PMC11640932/

[28] Humayun, M., Jhanjhi, N. Z., Niazi, M., Amsaad, F., & Masood, I. (2022). Securing drug distribution systems from tampering using blockchain. Electronics, 11(8), 1195.

[29] IBM. (n.d.). What is random forest? https://www.ibm.com/think/topics/random-forest

I. A. Shah, N. Jhanjhi and S. N. Brohi, "Proposing Model for Classification of Malicious SQLi Code Using Machine Learning Approach," 2024 1st International Conference on Innovative Engineering Sciences and Technological Research (ICIESTR), Muscat, Oman, 2024, pp. 1-6, doi: 10.1109/ICIESTR60916.2024.10798230.

[30] Internet Engineering Task Force. (2022, March 22). *Message Queuing Telemetry Transport (MQTT)-TLS profile for ACE framework*. https://www.ietf.org/archive/id/draft-ietf-ace-mqtt-tls-profile-17.html

[31] Kim, S., Lee, H., & Park, J. (2023). Hybrid framework for ransomware protection using effective decoy deployment and dynamic entropy check. Journal of Cybersecurity Research, 15(4), 112-135. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4496374

[32] Kosinski, M. (n.d.). *What is ransomware?* IBM. https://www.ibm.com/think/topics/ransomware

[33] Kritika, E. (2024). A comprehensive literature review on ransomware detection and defense mechanisms. *Computers & Security*, 120, Article 102994. https://www.sciencedirect.com/science/article/pii/S2772918424000444F

[34] Levacloud. (2024, July 26). *Evaluating CrowdStrike vs Microsoft Defender security*. https://levacloud.com/2024/07/26/crowdstrike-vs-microsoft-defender/

[35] Mahdi, Z. S. (2025). Advanced hybrid techniques for cyberattack detection on IoT networks. *Security and Privacy*, 2(1), Article e471. https://onlinelibrary.wiley.com/doi/full/10.1002/spy2.471

[36] Microsoft. (2023, December 31). What is zero trust architecture? Microsoft Security. https://www.microsoft.com/en-my/security/business/security-101/what-is-zero-trust-architecture

[37] Microsoft. (2025, April 22). *Prepare for a ransomware attack*. https://learn.microsoft.com/en-us/azure/security/fundamentals/ransomware-prepare

[38] Moradi, A., Zhang, Y., & Ghorbani, A. A. (2023). Hybrid framework for ransomware protection using effective decoy deployment and dynamic entropy check. *Journal of Information Security and Applications*, 68, 103197. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4496374

[39] NIST. (2020). Zero trust architecture (NIST Special Publication 800-207). National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf

[40] Page, C. (2023, August 25). *MOVEit, the biggest hack of the year, by the numbers*. TechCrunch. https://techcrunch.com/2023/08/25/moveit-mass-hack-by-the-numbers/

[41] Palo Alto Networks. (2019, December 31). What is zero trust architecture (ZTA)? https://www.paloaltonetworks.com/cyberpedia/what-is-a-zero-trust-architecture

[42] Palo Alto Networks. (2020, March 16). Microservices hybrid model. https://live.paloaltonetworks.com/t5/community-blogs/microservices-hybrid-model/ba-p/316677

[43] Park, D., & Dahger, G. (2025). Adaptive Honeypot Allocation in Multi-Attacker Networks via Bayesian Stackelberg Games. https://arxiv.org/abs/2505.16043

[44] Peris AI. (2025, March 16). *How AI and automation are enhancing ransomware detection*.

[45] https://peris.ai/post/how-ai-and-automation-are-enhancing-ransomware-detection

[46] Proofpoint. (2025, August 5). What is a software-defined perimeter (SDP)? https://www.proofpoint.com/au/threat-reference/software-defined-perimeter

[47] Ricoh USA. (2025, May 28). Ransomware containment and isolation. https://www.ricoh-usa.com/en/services-and-solutions/cloud-it-services/ricoh-cyber-security-services/ransomware-prevention-containment-and-isolation

[48] Sajid, M. S. I., Wei, J., & Al-Shaer, E. (2024). ranDecepter: Real-time identification and deterrence of ransomware attacks via active cyber deception. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(1), Article 22.

[49] https://arxiv.org/html/2508.00293v3

[50] Selvaraj, K. (2025). A conceptual framework to mitigate ransomware attacks on IoMT devices. *Procedia Computer Science*, 201, 450-457. https://www.sciencedirect.com/science/article/abs/pii/S1574013725000772

[51] Seetharam kakaraparthi, Durganjaneyulu immadisetty, & Maranco M. (2024). Enhanced honeypot security for intrusion detection and prevention systems using blockchain. *World*

Journal of Advanced Research and Reviews, *22*(1), 751–758. https://doi.org/10.30574/wjarr.2024.22.1.1065

[52] Saeed, S., Jhanjhi, N. Z., Khan, M. A., & Yadav, D. K. (2025). Digital transformation and cybersecurity challenges. Frontiers in Computer Science, 7, 1631362.

[53] Sulaiman, N. S., Nasir, A., Othman, W. R. W., Wahab, S. F. A., Aziz, N. S., Yacob, A., & Samsudin, N. (2021). Intrusion Detection System Techniques : A Review. *Journal of Physics: Conference Series*, *1874*(1), 012042. https://doi.org/10.1088/1742-6596/1874/1/012042

[54] Turton, W., & Mahtani, K. (2021, June 4). *Hackers breached Colonial Pipeline with one compromised password*. Al Jazeera. https://www.aljazeera.com/economy/2021/6/4/hackers-breached-colonial-pipeline-with-one-compromised-password

[55] Velosio. (2025, September 26). *CrowdStrike vs. Microsoft Defender for Endpoint*. https://www.velosio.com/blog/crowdstrike-vs-microsoft-defender-for-endpoint/

[56] Wikipedia. (2025, October 5). *Ransomware*. https://en.wikipedia.org/wiki/Ransomware

[57] Vigilant Asia. (2025, May 9). Zero trust architecture: Why it matters in 2025. https://www.vigilantasia.com.my/zero-trust-architecture/

[58] Zero Networks. (2025, May 26). *Ransomware protection strategies: Fixing 5 network security weaknesses*. https://zeronetworks.com/blog/ransomware-protection-strategies

[59] Riskhan, B., Roua, A., Mahaba, B. S., Uswa, D., Gheisari, M., & Sindiramutty, S. R. (2025). Enhancing Diagnostic Accuracy for Breast Cancer Using Classical-Quantum Hybrid and Transfer Learning Technique. *Journal of Soft Computing and Data Mining, 6*(2), 41–56. https://penerbit.uthm.edu.my/ojs/index.php/jscdm/article/view/22061

[60] Keon, B. T. W., Zhe, L. C., Cahyono, R. C., Balakrishnan, S., Sindiramutty, S. R., Wei, G. W., & Hendrawati, T. D. (2025). Enhancing Trust in EEG-based Depression Classification: A LIME-powered XAI Approach to Dissecting Deep Learning Black Box Results. *2025 International Conference on Metaverse and Current Trends in Computing (ICMCTC)*, 1–6. https://doi.org/10.1109/icmctc62214.2025.11196423

[61] Kiyani, F. F., Hamid, B., Humayun, M., Sindiramutty, S. R. a. L., & Chowdhury, S. (2024). Discovery of Influential Publications Using Research Article's Usage Context. *2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC*, 1–7. https://doi.org/10.1109/etncc63262.2024.10767576

[62] Krishnan, S., Thangaveloo, R., Rahman, S. B. A., & Sindiramutty, S. R. (2021). Smart Ambulance Traffic Control system. *Trends in Undergraduate Research*, *4*(1), c28-34. https://doi.org/10.33736/tur.2831.2021

[63] Linqiang, Y., Sindiramutty, S. R. a. L., Ashraf, H., Muzammal, S. M., Balakrishnan, S. a. P., Gupta, S., & Kavita, N. (2024). Intelligent Household Waste Classification System Based on Machine Learning. *2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, 760–768. https://doi.org/10.1109/etncc63262.2024.10767563

[64] Lim, M., & Abdullah, A. Jhanjhi NZ Performance optimization of criminal network hidden link prediction model with deep reinforcement learning (2021) Journal of King Saud University-Computer and Information Sciences, 33(10), pp. 1202- 1210.

[65] Sanjaya, F. N., Balakrishnan, S., Sindiramutty, S. R., & Narputro, P. (2025). XAI Technology in EEG Signal Based Disease Detection Models. *2025 International Conference on Metaverse*

and Current Trends in Computing (ICMCTC), 1–7. https://doi.org/10.1109/icmctc62214.2025.11196271

[66] Sindiramutty, S. R., Jhanjhi, N., Tan, C. E., Lau, S. P., Muniandy, L., Gharib, A. H., Ashraf, H., & Murugesan, R. K. (2024a). Industry 4.0. In *Advances in logistics, operations, and management science book series* (pp. 342–405). https://doi.org/10.4018/979-8-3693-1363-3.ch013

[67] Sindiramutty, S. R., Jhanjhi, N. Z., Tan, C. E., Tee, W. J., Lau, S. P., Jazri, H., Ray, S. K., & Zaheer, M. A. (2024b). IoT and AI-Based Smart Solutions for the Agriculture Industry. In *Advances in computational intelligence and robotics book series* (pp. 317–351). https://doi.org/10.4018/978-1-6684-6361-1.ch012

[68] Sindiramutty, S. R., Prabagaran, K. R. V., Akbar, R., Hussain, M., & Malik, N. A. (2024a). Generative AI for Secure User Interface (UI) design. In *Advances in information security, privacy, and ethics book series* (pp. 333–394). https://doi.org/10.4018/979-8-3693-5415-5.ch010

[69] Sindiramutty, S. R., Prabagaran, K. R. V., Akbar, R., Hussain, M., & Malik, N. A. (2024b). Overview of Generative AI techniques for cybersecurity. In *Advances in information security, privacy, and ethics book series* (pp. 1–52). https://doi.org/10.4018/979-8-3693-5415-5.ch001

[70] Sindiramutty, S. R., Prabagaran, K. R. V., Jhanjhi, N. Z., Ghazanfar, M. A., Malik, N. A., & Soomro, T. R. (2024c). Security Considerations in Generative AI for web Applications. In *Advances in information security, privacy, and ethics book series* (pp. 281–332). https://doi.org/10.4018/979-8-3693-5415-5.ch009

[71] Sindiramutty, S. R., Prabagaran, K. R. V., Jhanjhi, N. Z., Murugesan, R. K., Brohi, S. N., & Wei, G. W. (2024d). Generative AI for threat intelligence and information sharing. In *Advances in digital crime, forensics, and cyber terrorism book series* (pp. 191–234). https://doi.org/10.4018/979-8-3693-8944-7.ch006

[72] Sindiramutty, S. R., Prabagaran, K. R. V., Jhanjhi, N. Z., Murugesan, R. K., Malik, N. A., & Hussain, M. (2024e). Ethics and transparency in secure web model generation. In *Advances in information security, privacy, and ethics book series* (pp. 411–464). https://doi.org/10.4018/979-8-3693-5415-5.ch012

[73] Sindiramutty, S. R., Tan, C. E., & Wei, G. W. (2024). Eyes in the sky. In *Advances in information security, privacy, and ethics book series* (pp. 405–451). https://doi.org/10.4018/979-8-3693-0774-8.ch017

[74] Weiqi, X., Hooi, S. T. C., Sindiramutty, S. R. a. L., Asirvatham, D. a. L., Kumar, D., & Verma, S. (2024). Surface Anomaly Detection Using Machine Learning Technique. *2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, 1–7. https://doi.org/10.1109/etncc63262.2024.10767562

[75] Ying, X., Murugesan, R. K., Sindiramutty, S. R., Wei, G. W., Balakrishnan, S., Kumar, D., & Verma, S. (2024). Scene Text Recognition using Deep Learning Techniques. *2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, 1–9. https://doi.org/10.1109/etncc63262.2024.10767484