iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VI **Month of publication:** June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83649

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

Intelligent Ransomware Defense System with Digital Forensics Integration

Sohel Shaikh[1], Shubham Kokane[2], Vikas Khandekar[3], Yugandhar Dorugade[4], Dr. Vina Lomte[5]
[1, 2, 3, 4]*Department of Information Technology Dr. D. Y. Patil Institute of Technology, Pune, India*
[5]*Project Guide, Department of Information Technology Dr. D. Y. Patil Institute of Technology, Pune, India*

Abstract: ***Ransomware has been a major concern in the domain of cyber security insofar as it has caused losses estimated in billions of dollars across the globe. The traditional means of guarding against it, that are based on signatures, tend to be less effective in cases of modern ransomware that employing polymorphism and zero-day attacks very effectively. Therefore, there has been a paradigm shift in the detection process that has achieved a high level of accuracy in behavior-based, entropy-based, and machine-learning techniques but faces a major drawback in a lack of transparency in decision-making.***
This paper examines the prevailing approach to the detection of ransomware, and the trade-off that needs to be made between the accuracy and explainability of the approach. To address the gap in the literature, this paper describes the development of a light-weight host-based approach to the detection of ransomware, utilizing real-time entropy-based monitoring, and Machine Learning. Additionally, this paper includes the use of Explainable AI techniques, specifically SHAP and LIME, to offer explanations for each ransomware detection event. It is hoped that the new solution developed in this paper shall guarantee high accuracy and real-time detection, while remaining transparent for forensic analysis.

I. INTRODUCTION

Ransomware has turned into a ‘silent’ pandemic that ‘quietly’ breaks into systems, ‘steals data,’ and holds it ‘hostage’ until ‘ransom payments’ are received. It has grown from a ‘petty crime’ to ‘a significant’ level of the ‘cyber security threat.’ As of 2023, the cost of this pandemic has been estimated to be a staggering $1.1 billion, a 140% increase from the last year, and ransomware attacks accounted for seven out of ten cyber attacks. Apart from the statistics, the cost is personal, and this is where downtime, reputation, and the process of recovery come into play. As threats are becoming more sophisticated, it is the need of the hour that businesses pay attention to protection, detection, and response. As sophistication is increasing, it is the need of the hour that the business pays attention to the concepts of protection, detection, or response.

However, recent events, like the data breach that took place in 2024 in Conduent Business Services, have shown the disastrous effects of such failures. The data breach involved the personal data of more than 10.5 million individuals, and the current average cost of recovery per incident, not including ransom payments, is $1.5 million.

This is due to the existence of a weakness in the traditional defensive system. The traditional Antivirus software, which is the backbone of the traditional cyber security system, is almost entirely dependent on signatures. This implies that the software relies on a database of known malware IDs, such as file hashes or code bits, and will only flag a file as malicious if it is known. This type of reactive approach is extremely inefficient in today’s attacks. In fact, today’s attackers use zero-day exploits, which lack a patch or a signature, and polymorphic malware, which has the capability to change its code to generate a large number of new and unique signatures for each attack. This implies that such malware can evade about 99% of signature-based solutions. Moreover, the development of fileless malware, which carries out all its operations in memory without writing a malicious file to disk, has made the signature matching of files unnecessary.

The inefficiency of signature-based antivirus software has created a need for a paradigm shift for the defenders. They are no longer interested in knowing "What is this file?" but are now interested in knowing "What does this process do." This has created a need for more intelligent and capable detection software, whether it is behavior-based detection software, anomaly-based detection software, or ML models. These models create a known system activity baseline and raise an alert on system activity that is deemed to be abnormal, such as a user-space process that reads, encrypts, and deletes thousands of files.

These complex systems have brought about a new and serious "black box" problem to the table in detection problems. A machine learning system may be able to accurately determine a process as malicious, but it will typically give no more than a score or a flag, not an explanation. This lack of explanation is a problem for the operations side of things, but it is not acceptable for the forensic side of things. A digital forensic analyst cannot give "the model said so" as evidence in court.

It would be dangerous for a security operations analyst to begin an expensive incident response action, such as turning off a critical server, based solely on the receipt of an alert if they do not understand why the alert was generated; it could be an expensive false alarm. The lack of clarity causes an “accountability gap” that makes it difficult to make decisions and renders the data unusable as evidence.

However, in order to resolve this issue, we must have Explainable AI, or XAI. XAI uses methods that offer clear explanations for the results obtained by complex AI systems. In this manner, a human can trust the system, staying within legal and moral parameters. This paper will begin by examining ransomware detection systems in order to show the huge difference in explainability. It will then describe the problem and propose a new design that incorporates XAI as a trustworthy and forensically accurate solution.

II. LITERATURE REVIEW

This review will discuss the evolution of ransomware detection, from traditional approaches to modern intelligent systems. We will analyze these approaches in terms of their effectiveness in threat detection, data collection, performance in real-world scenarios, and, most importantly, their explainability.

A. Survey of Ransomware Detection Methodologies

The approach that was being taken in the early days of malware protection was basically solely focused on finding known signatures. Antivirus programs were searching for known code patterns, or digital signatures, that were linked to known malware that we were afraid of. This technology was doing all right for a long time, especially when it came to more well-known types of malware. However, it soon became apparent that, as more advanced types of ransomware attacks began to emerge, the weaknesses of this technology were beginning to show. More advanced versions of this type of malware use polymorphism, which is constantly changing and combining their code, while others use zero-day attacks that have not yet left any signatures.

As the agility of ransomware escalated, there was a transition towards detection based on tracing program behavior. For instance, ShieldFS traces the very small backstage file system activity involving reading, writing, renaming, deleting, and searching for typical patterns of stinky code running in the background. ShieldFS could not perform better than a 97.7% hit rate, but this was accompanied by very noticeable slowdowns and vulnerability to divide attacks. The Crypto Drop is another example of behavior-tracing code watchdogs, which rely on a much larger set of characteristics, including sudden changes in file types, a sudden increase in entropy, and fishy directory traversal patterns. The Crypto Drop claimed flawless performance in tests, but this code was prone to producing false positives in large quantities; more importantly, however, it would raise the alarm only when ten or more files had been infected, on average.

This was succeeded by light-weight designs that concentrated on entropy, aiming at the interception of high entropy write operations, which are prevalent in encryption algorithms. For instance, take the design of Rcryptect, which employed statistical testing in addition to block-level entropy analysis. It worked well with a negligible performance overhead. However, the design is susceptible to evasion attacks such as entropy sharing and staged encryption, which could emerge as a result of the concentration on entropy analysis. The current state of the art indicates that machine learning designs have already attained the detection rate of high 90s, often breaking the 95% threshold and ready to cross the 99% barrier. Conventional models (Random Forest, SVM, and XGBoost) and deep learning models have been trained on thousands of labeled malware and benign samples.

Despite their good performance, all these models are very difficult to interpret because of their complexity. They function like "black boxes" in reality because they give a verdict without any explanations or reasons. Because of this, the lack of transparency in these models makes them unsuitable for forensic purposes because all alerts must be justified.

B. Analysis of Monitoring Layers and Attributes

One of the most critical areas of research in the detection of ransomware is the extent of system activity that we are monitoring. The extent of monitoring at the kernel level is the highest, and this is what ShieldFS and UNVEIL are based on. ShieldFS was functional within the Windows Minifilter Driver model, monitoring IRPs or I/O Request Packets. Prior to these occurring in the file system, it monitored file type, write entropy, and directory access patterns. UNVEIL also monitored filesystem activity and user interface tampering patterns of ransomware.

While kernel-level visibility can help in the early detection of threats with high accuracy, it has downsides: hard to implement, more system overhead, and may cause system instability. Even a single mistake made in a kernel driver can crash an entire OS, and thus is not too suitable for real business environments.

To overcome these challenges, relatively more recent approaches such as Crypto Drop and Rcryptect have been based on user-space monitoring. Crypto Drop does all its monitoring in the user space, monitoring the amount of entropy change, the difference between read and write entropy, deletions of files, and directory traversal patterns. Rcryptect uses lightweight block-level checks to monitor the encryption process in real-time. In this paper, the system will continue in this fashion using Python-based inotify/watchdog techniques to monitor file writes and modifications, calculate the entropy of each file, and use these as inputs to a machine learning model that provides explainable results.

This is a definite trade-off in engineering, and the monitoring in the user space gives up some of the low-level visibility in favor of safety, portability, and ease of deployment. This is a very important aspect of how well it can be adopted in the enterprise and tested on different platforms.

C. Data Acquisition and Feature Engineering in Detection Systems

In fact, the literature on the detection of ransomware is also very much dependent on the processing of raw system data into meaningful features. Other research works also involved large kernel-level IRP capture tools such as IRPLogger, which logged over 1.7 billion I/O packets from more than 2,000 safe applications. These were used to identify the normal patterns of I/O and extract statistical and heuristic features such as file type usage, rename patterns, and entropy.

The entropy-based approaches confirmed that Shannon entropy and entropy change were the sure indicators of file encryption. The study conducted on a large number of user files and ransomware samples confirmed that the difference between read entropy and write entropy was a reliable indicator of malicious encryption. Further improvements included the concept of a sliding window and block entropy analysis, thus allowing real-time detection even when the file is being partially encrypted.

The improved system improves this by using statistical and timing features derived from safe and malicious write patterns using lightweight user-space file system monitoring. These features are then used in the ML model, along with techniques such as SHAP or LIME, to provide accuracy as well as explainability.

D. Performance Evaluation of Detection Models

In the literature, the accuracy of detection of the contemporary ransomware detection tools is always high. ShieldFS achieved a 97.7% detection rate on 305 new samples of ransomware and still provided the user with different options for recovery. Crypto Drop claimed a 100% detection rate with no false positives in its test set of 15 real families of ransomware. However, the most important weakness is that the threat is normally detected only after a large amount of the user's data has been encrypted.

Among these, machine learning classifiers such as XGBoost have been found to provide accuracy at the level of 95%-99% and have even surpassed rule- and statistic-based systems. Deep learning systems that were tested using more than 10,000 data samples have provided accuracy at the level of 99.2%. All these parameters confirm that accuracy levels of 99% are technically possible in detection systems.

However, it is not sufficient to have high accuracy. One of the well-known issues is that of the so-called "day-after problem," where the system is capable of identifying an attack due to ransomware, but only after "irreparable harm has been wrought." What is required is real-time or near real-time detection that will prevent malicious processes from actually beginning to encrypt the data. This is where the proposed solution will fill this gap.

E. Explainability and Forensic Utility

The key enabler in the progress of ransomware detection is the way the pieces of the puzzle work together to offer accuracy and transparency in the forensic process. Conventional signature-based approaches are very easy to understand since there is tangible evidence available-a hash or pattern match, which can be justified if needed in a court of law. These approaches are terrible, however, at polymorphic ransomware.

Heuristic software such as Crypto Drop is quite easy to comprehend when compared to more complex software. It looks for very visible warning flags such as a sudden spike in the entropy of files or a very peculiar pattern of files that are not accessed very often, and then alerts when these flags are reached at a particular level. It is quite easy to understand why a particular file was flagged as being suspicious.

Meanwhile, the modern methods of machine learning and deep learning provide very high accuracy but do not provide any meaningful or explainable reasons for their decisions. Mostly, they only provide a classification score rather than any motives that can underlie them. Thus, such methods are limited as forensic evidence or as reliable tools in cybersecurity. Vague alerts, for instance, cannot be acted on by courts and incident response teams.

However, the lack of explainability in black-box models can be overcome by the application of SHAP and LIME, which will ensure that machine learning decisions are more understandable to humans. The first models prove that XAI-boosted detection can restore the forensic value that black-box models have lost.

The proposed system is directly addressing this issue by integrating machine learning detection with XAI-based explanation so that each alert is accompanied by an explanation that is understandable.

F. Summary of Literature

It is clear from the literature that although ML-based detection is a highly necessary and powerful tool, the "black box" problem of ML-based detection is a dead end for digital forensics. The future is not to deny ML but to improve it. Techniques such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (Shapley Additive exPlanations), which can explain the predictions of the models based on features, are the only hope. This review proves that there is no existing system that is integrated, lightweight, and real-time, capable of meeting the requirement of high-accuracy detection as well as the essential requirement of forensic explainability.

III. GAP IDENTIFICATION AND PROBLEM STATEMENT

Section II explores the literature in this area and highlights the obvious insight that has been gained from the forensic explainability literature review; although various components of the effective defense strategy have been developed in the past, such as entropy-based monitoring methods and high-accuracy machine learning classifiers, they have never been combined together in a single, coherent, and forensically defensible manner.

The signature-based antivirus engines are capable of providing a correct explanation for their detection as the detection is explicitly triggered by hash or pattern matching. However, the antivirus engines are not capable of detecting zero-day, polymorphic, and fileless ransomware attacks, thus proving inefficient in the presence of modern ransomware threats. The behavior-based anomaly detectors are more robust against unknown ransomware attacks but are usually accompanied by high false-positive rates. They also lack information regarding the forensic explanation required in post-incident analysis.

The entropy-based detection techniques have shown high predictability in the detection of encryption behaviors; however, most of these techniques are standalone modules and do not interact with real-time alerting systems that could provide readable explanations in the human language.

Machine learning-based detection systems have the highest accuracy rate from the current literature, which is always above 95 to 99 percent. They also bring a completely new set of problems because there is no transparency in the decision-making process in the AI system. They are like black boxes that make decisions, and none of the decisions and how they are made are traceable.

The combination of all these problems indicates the existence of an outstanding gap. There has not been a system that could provide real-time host-level ransomware detection and provide, for each alert, simple human-understandable explanations with low computational complexity.

This gap directly leads to the inspiration of the proposed system, which aims to combine the detection accuracy of machine learning systems with the forensic transparency needed for proper and defensible cybersecurity operations.

A. Problem Statement

On the basis of this gap, the problem that the research study aims to solve is: “The current methods for real-time detection of ransomware either fail to detect new attacks or produce black-box alerts that forensic analysts cannot understand. There is a need for the development of light-weight, host-based systems that can monitor file operations, identify high-entropy writes, and provide XAI-supported explanations for each alert.”

IV. PROPOSED SYSTEM ARCHITECTURE

We propose a new lightweight framework that combines real-time anomaly detection and post-alert XAI explanation to address the problem that has been identified. Our proposed framework will try to address the “Forensic V-Shape” paradox by combining the puzzle pieces that have been validated through the literature review.

A. High-Level Architecture

It has four major modules that run in the user space of the host OS. Regarding the details in Table 02, the design focuses on low overhead, safety, and ease of implementation. The system architecture is shown in the diagram below:

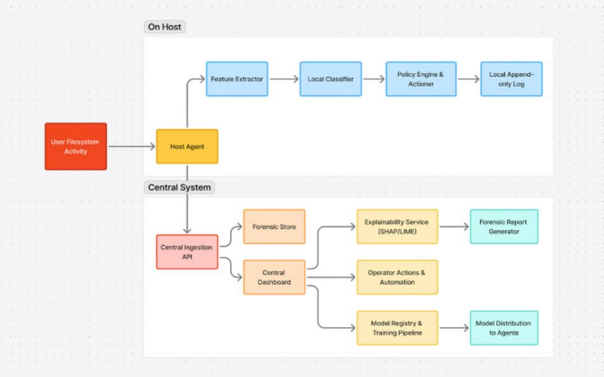


Fig.1 System Architecture of Proposed Ransomware Detection Framework

B. System Flow Explanation

The operational flow is set up for real-time detection and later explanation, ensuring speed and accountability.

1) The File System Monitor captures all file write events on the host.

2) For each type of event it monitors, the Anomaly Detection Engine computes the Shannon entropy of the data being written and how unusual or random it looks, and then based on this data, as well as other factors such as how quickly the file is being modified and whether the process has a good or bad reputation, makes a determination of whether it looks suspicious by passing all of this data to a small machine learning model.

3) If the classifier marks the event as unusual (exceeding a confidence threshold), the event goes to the Causal Logger. This module securely records all event details PID, process name, file path, entropy delta, and timestamp in an unchangeable, append-only JSON log for chain of custody. The "why" (the explanation) is then exported for forensic review or dashboard visualization. Explainer module.

4) The XAI component uses a framework, such as SHAP, to generate an explanation for the decision reached by the ML component. It generates an explanation that can be interpreted by a human. For example: The system determined the event to be 95% malicious most of that score (70%) comes from a sharp rise in entropy, while another 25% is due to an unusually fast rate of file writes.

5) This alert is completely supported and offers both the result of the detection ("what happened") and the reason for the detection ("why it happened"). This alert is then sent for forensic analysis or displayed on a monitoring dashboard.

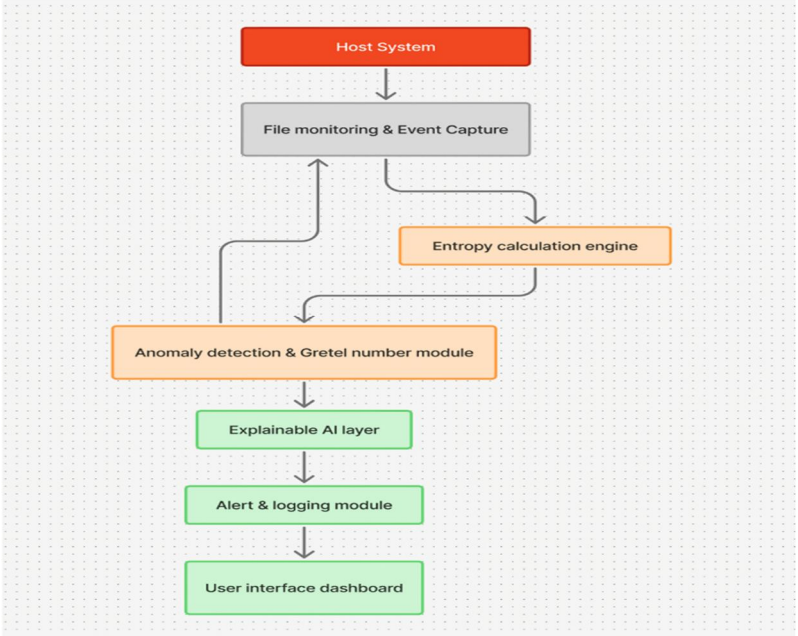


Fig. 2. Host-Based Detection and Centralized XAI Framework

C. Synthesis and Main New Ideas

It is not a question of examining each piece separately, but rather how all the pieces of the puzzle fit together. The goal is to solve the "Forensic V Shape" puzzle as explained in Table 05 by integrating the sharp and reliable information obtained from the entropy analysis with the reliable deductions made by XAI.

The theory relies on two large concepts: the possibility of detection on the host level in a lightweight user-space environment, as explained in the deployment trade-offs in Table 02, and the integration with XAI to produce forensically sound alerts, solving the "black box" problem as explained in most modern systems today based on ML

Extensibility: The system will function flawlessly in the VM/test mode and in the enterprise without depending on kernel dependencies. Looking ahead into the future, there are visions of the framework functioning in other environments, perhaps even IoT or cloud.

D. Sequence Diagram Explanation

The above diagram shows, in a simplified way, the process of handling the file event in the system. The File Watcher logs the file event, and the Processing Worker extracts the critical information. The Anomaly Engine relies on the information available in the machine learning model to determine if the activity is anomalous. If it is, then the XAI module will explain the activity. The Response Executor will then notify or respond to the files.

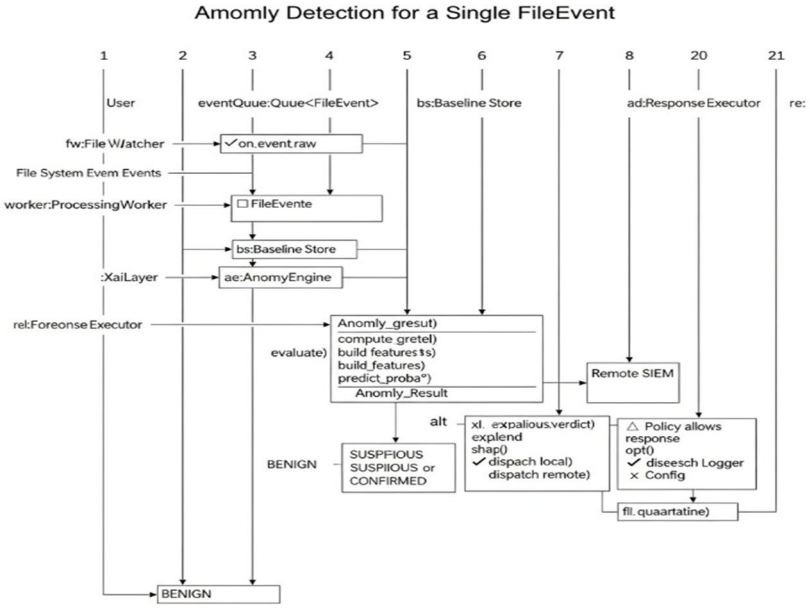


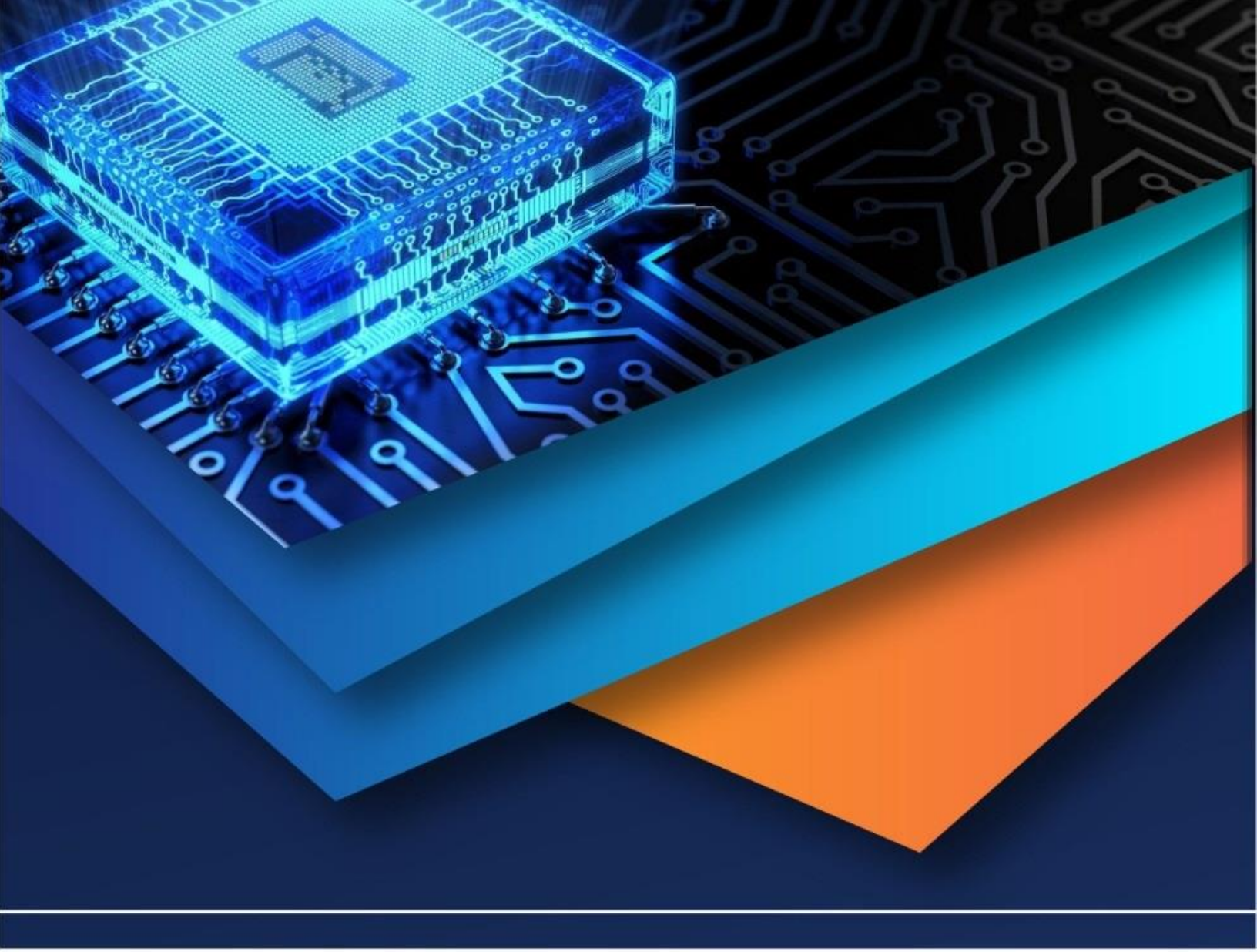
Fig. 3. Sequence Diagram for Anomaly Detection of a Single File Event

V. CONCLUSION

The paper traces the history of ransomware detection, a costly and desperate game. We are now past the old and signature-based approaches to smarter and more correct machine learning approaches. But our research indicates that some of the consequences of the quest of accuracy is that it leads to the establishment of an accountability gap. The current state of systems is that of black boxes that are not easily trusted to do forensic analysis, their behavior is opaque, and this is the biggest issue in legal and critical circumstances. We demonstrate by multi-table analysis (Tables 01-05) that the community is split between simple and understandable yet not accurate enough models and accurate yet complicated to understand models. This issue is known as the Forensic V-Shape and it is the greatest issue in the present condition of the cybersecurity world. This issue can be resolved by the proposed system. We suggest a system that is accurate and accountable with the assistance of numerous studies about the detection engine that is lightweight and real-time entropy-based and multiple existing explainable AI tools like SHAP and LIME. It does not only raise an alarm but also justifies the security operators and one they can rely on and even prove in the court. Malware detection and description of the malware are part of a vital shift between the process of detecting the malware and providing security intelligence which is credible, actionable, and defensible.

REFERENCES

[1] A. Al-rimy, B. A. S., et al., "Advanced Machine Learning for Ransomware Detection," 2024.
[2] BleepingComputer, "BPO giant Conduent confirms data breach impacts 10.5 million people," 2025.
[3] CrowdStrike, "AI Decision-Making with SHAP: A Game-Theoretic Approach to Explainable AI in Cybersecurity," 2024.
[4] CrowdStrike, "Benefits of Machine Learning in Cybersecurity," 2024.
[5] Continella, A., et al., "ShieldFS: A self-healing, ransomware-aware filesystem," in Proc. ACSAC, 2016.
[6] Continella, A., "ShieldFS: The Last Word in Ransomware-Resilient Filesystems," Black Hat USA, 2017.
[7] Fortinet, "Ransomware Statistics: What to Know in 2024," 2024.
[8] GCA, "How Ransomware Can Evade Antivirus Software," 2023.
[9] Huang, C., et al., "Explainable Machine Learning for Cyber Threats: A Survey," arXiv preprint arXiv:2208.14937, 2022.
[10] IBM X-Force, 2025 Threat Intelligence Index, IBM Corporation, 2025.
[11] Infosecurity Magazine, "Conduent Data Breach Impacts Over 10 Million Individuals," 2025.
[12] Kharraz, A., Arshad, S., Mulliner, C., Robertson, W., and Kirda, E., "UNVEIL: A Large-Scale, Automated Approach to Detecting Ransomware," in Proc. USENIX Security, 2016.
[13] Kharraz, A., "Overview of the design of the I/O access monitor in UNVEIL," 2016.
[14] Kirda, E., "UNVEIL: A large-scale, automated approach to detecting ransomware (keynote)," 2016.
[15] Lundberg, S. M., and Lee, S. I., "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems (NeurIPS), 2017.
[16] Medium, "Explainable AI in Cybersecurity: Ensuring Transparency," 2024.
[17] MDPI, "A Comparative Analysis of SHAP and LIME for Forensic Cybersecurity," 2024.
[18] Palo Alto Networks, "Advanced Endpoint Security vs Antivirus," 2024.
[19] Palo Alto Networks, "Explainable AI (XAI) in Cybersecurity," 2024.
[20] Park, J., and Kim, J., "Rcryptect: Real-Time Detection of Cryptographic Function in the User-Space Filesystem," ETRI Journal, 2019.
[21] Ribeiro, M. T., Singh, S., and Guestrin, C., "Why Should I Trust You?: Explaining the Predictions of Any Classifier," in Proc. ACM SIGKDD, 2016.
[22] Sangfor Technologies, "Machine Learning in Cybersecurity: Benefits and Challenges," 2024.
[23] Scaife, N., Carter, H., Traynor, P., and Butler, K. R. B., "CryptoDrop: An Early-Warning Detection System for Ransomware," in Proc. IEEE ICDCS, 2016.
[24] SentinelOne, "Signature-Based vs. Behavioural AI Detection," 2024.
[25] Sentinel-Overwatch, "What Are the Limitations of Signature-Based Intrusion Detection?" 2024.
[26] SETS India, "A Review on Explainable AI for Cybersecurity," 2024.
[27] Sharma, A., and Gupta, S., "The Role of Explainable AI (XAI) in Forensic Investigations: Enhancing Trust and Transparency," Journal of Forensic Sciences and Digital Investigation, 2025.
[28] Sophos, The State of Ransomware 2025, Sophos Ltd., 2025.
[29] Spin.ai, "Ransomware Detection Using Machine Learning," 2024.
[30] The Data Scientist, "Explainable AI: Making Cybersecurity Clear and Trustworthy," 2024.
[31] Zhang, X., et al., "Ransomware Detection Using Machine Learning," 2023.
[32] Zhang, Y., et al., "Real-Time Ransomware Detection Using Entropy Analysis," IEEE Trans. Inf. Forensics Security, 2019.
[33] Zhang, Y., "Accuracy and performance of randomness tests in distinguishing encrypted from non-encrypted files," 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY