Shared Defense Response in Corporate Networks

Stefan Tafkov[1,3][0009-0004-9239-3098], and Zlatogor Minchev[1,2,3][0000-0003-2479-5496]

[1]Institute of ICT, Bulgarian Academy of Sciences,
Acad. Georgi Bonchev Str., Bl. 25A, Sofia, 1113, Bulgaria
[2]Institute of Mathematics & Informatics, Bulgarian Academy of Sciences,
Acad. Georgi Bonchev Str., Bl. 8, Sofia, 1113, Bulgaria
[3]Centre for Implementation of Scientific Research on Digitisation of the Economy in an Environment of Big Data (DEEBD), Sofia, Bulgaria
stefan.tafkov@iict.bas.bg, zlatogor@bas.bg

Abstract. In recent years, malware and particularly ransomware has evolved rapidly, employing increasingly sophisticated techniques to infiltrate and compromise modern computing environments. This escalation highlights the urgent necessity for adaptive endpoint defense mechanisms capable of detecting and mitigating emerging threats in real time. The present work introduces a Machine Learning-based Endpoint Detection and Response (ML EDR) model designed to provide dynamic, behavior-driven protection against both malware and ransomware attacks. The proposed system leverages multi-layer telemetry collected from distributed endpoint sensors used to model system behavior, identify anomalies, and predict malicious actions before they fully execute. By analyzing traffic flows, process activity, and file operations in a dynamic sandboxed environment, the model learns behavioral signatures directly from infected samples. These signatures enable the system to classify threats and anticipate future malicious steps with high accuracy. Integration with a Cloud Intelligence Network and cloud-assisted file analysis enhances the model's adaptability, enabling rapid updates, collaborative threat intelligence sharing, and large-scale pattern correlation. The multi-layer monitoring architecture ensures continuous visibility across endpoints, enabling early-stage detection of polymorphic and zero-day ransomware variants. Experimental results demonstrate that the adaptive ML-based EDR model improves detection precision and significantly reduces response time to emerging threats. The study contributes to a scalable, self-evolving defense mechanism suitable for modern enterprise security ecosystems.

Keywords: Ransomware, Machine Learning, EDR (Endpoint Detection and Response), Behavioral Analysis, Telemetry Detection, Neural Network/Residual Neural Network (ResNet), Threat Intelligence, Malware Cloud Intelligence, Cloud-Based Analysis.

1 Introduction

The proliferation of ransomware over the past decade has transformed it from a relatively isolated threat into one of the most disruptive and economically damaging forms

Research Paper
DOI: https://doi.org/10.46793/BISEC25.110T
Part of ISBN: 978-86-89755-40-4

of cyberattack [1], [3]. Modern ransomware campaigns employ polymorphism, fileless techniques, lateral movement, and multi-stage attack chains that significantly complicate traditional signature-based defenses [4].

As adversaries refine their tradecraft, endpoint security solutions must evolve toward adaptive, intelligence-driven architectures capable of analyzing behavioral indicators in real time and responding autonomously to unknown attack variants.

To address these challenges, the present study investigates the development of a Machine Learning-based Endpoint Detection and Response (EDR) model engineered specifically to counter advanced malware and ransomware behavior.

The model integrates deep learning techniques [2], multi-source telemetry, and cloud-assisted intelligence to dynamically characterize malicious activity at the network and system level [6], [7]. By correlating anomalous process execution, filesystem mutations, registry modifications, and command-and-control (C2) communication patterns, the system aims to detect early-stage ransomware activity prior to the encryption phase.

A central focus of this work is the analysis of behavioral artifacts generated during live ransomware execution. Controlled experiments were performed within an isolated sandbox to capture detailed telemetry from several prominent ransomware families, including but not limited to: REvil, WannaCry, Loki, TeslaCrypt, Petya, CryptoLocker, and CrySIS. These samples were selected due to their diverse infection strategies, encryption routines [11], and C2 communication protocols, providing a heterogeneous dataset for model training and validation. Observing how each variant acquires environmental context, establishes persistence, negotiates encryption keys, and communicates with remote servers enables the construction of an enriched behavioral dataset suitable for deep neural network training.

Building upon this dataset, the proposed EDR architecture employs a hybrid static-dynamic analysis pipeline supported by residual neural networks (ResNet) [5]. Static analysis provides rapid pre-execution filtering and feature extraction, while dynamic behavioral monitoring ensures high-fidelity detection of zero-day and morphing ransomware strains. Complementing this, cloud-based intelligence modules aggregate anonymized telemetry from distributed sensors, enhancing pattern recognition and enabling real-time distribution of newly learned threat indicators.

The introduced system further explores mechanisms for intercepting and analyzing ransomware communication flows with the objective of extracting cryptographic metadata and understanding key-exchange procedures. This capability supports proactive containment and lays the groundwork for potential partial data restoration in specific attack scenarios.

The rest of the paper presents the architectural framework, machine learning methodology, feature engineering approach, and experimental evaluation. Results demonstrate the feasibility of an adaptive, ML-driven EDR model capable of identifying diverse ransomware behaviors with high accuracy and reduced response latency, contributing a scalable defense mechanism for modern enterprise environments.

2 Methodology

The methodology applied in this study follows an integrated scientific and engineering workflow that combines dataset construction, feature transformation, machine learning model development, EDR orchestration logic, and experimental validation inside controlled and real-world execution environments. All phases are interconnected to ensure that the resulting defense system is capable of detecting, interpreting, and mitigating both known and unknown ransomware and malware variants in an adaptive manner.

The research process begins with the acquisition and preparation of a comprehensive dataset representing diverse categories of malicious and benign software. Ransomware samples from more than twenty families, including both legacy and modern polymorphic variants, were collected from public repositories, industry intelligence feeds, and internally operated honeypots. These samples were executed in a controlled sandbox laboratory configured with kernel-level instrumentation.

The aim was to observe and record all measurable behaviors, such as file system bursts, encryption attempts, suspicious memory allocation patterns, crypto-API usage, registry persistence routines, process forking structures, lateral movement indicators, and network communications with emulated or real command-and-control infrastructures [9], [10], [12]. Simultaneously, benign software was executed under identical conditions to capture normal operating baselines, ensuring sufficient contrast for machine learning–based discrimination and for controlling false-positive rates. Following data acquisition, the raw static and dynamic telemetry was preprocessed into structured feature representations. Static features extracted from the PE files included section-level entropy, header fields, opcode n-grams, fuzzy hashes, library call imports and exports, string embeddings, and partial control-flow approximations.

Dynamic telemetry was transformed into behavioral sequences and graphs built from event streams such as file modifications, registry writes, inter-process communication events, and memory operations. Each behavioral series was normalized, time-aligned, and encoded into metadata vectors suitable for neural network ingestion. The system also incorporated controlled honeypot artifacts placed in high-value directories. These bait files, when modified, provided high-confidence behavioral ground truth, enabling the machine learning models to learn reliable patterns strongly correlated with ransomware encryption phases.

All extracted features were then used to train a multi-layer machine learning architecture embedded in the EDR system. The first (primary) layer consists of a lightweight classifier designed to rapidly assess incoming samples and eliminate clearly benign activity. This ensures low latency and reduces the computational burden on deeper models.

The second and primary layer is a deep residual neural network (DRNN) tailored to interpret both static and dynamic representations of unknown executables. The DRNN receives embedded PE structural features, opcode distributions, event-sequence matrices, and graph-derived behavioral signatures, enabling it to classify threats even when obfuscation or polymorphism is present.

To support resilience and reduce false positives, the DRNN incorporates calibrated confidence scoring, adversarial training with benign-noise augmentations, and

uncertainty quantification, allowing the EDR agent to detect ambiguous classifications and request additional validation from the cloud intelligence service. Parallel to the neural network analysis, the EDR agent performs independent real-time behavioral correlation. Each process running on the endpoint is continuously monitored, and its actions are integrated into a dynamically updated graph representing file interactions, registry operations, memory writes, thread spawns, DLL injections, and network flows.

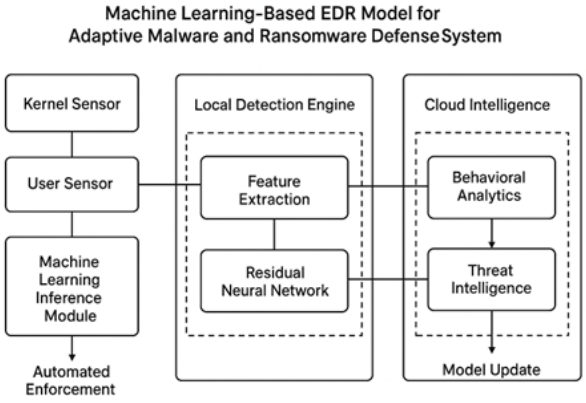

Fig. 1. Architecture of the Machine Learning-Based EDR system illustrating kernel and user-level sensors, the local detection engine with feature extraction and Residual Neural Network modules, and the cloud intelligence layer responsible for behavioral analytics, threat intelligence generation, and continuous model updates.

This behavioral map is compared to previously observed ransomware patterns, enabling the system to detect attacks even before the neural network completes its inference cycle. When the behavioral model identifies high-risk sequences (such as rapid recursive file encryption, unauthorized manipulation of shadow copies, or consistent API patterns related to cryptographic routines) the EDR agent can suspend the suspect process at the kernel level.

The suspended process is then pushed through the deeper ML layers and cloud validation routines before a final decision is made. This suspension-and-verify methodology is crucial for minimizing false alarms while ensuring timely intervention during real attacks.

To further enhance detection accuracy and incident response, the methodology integrates cloud-based intelligence throughout the workflow. When the local EDR agent encounters ambiguous behavior or low-confidence ML predictions, a secure channel transmits anonymized metadata, fuzzy hashes, encryption artifacts, and behavioral vectors to the Cloud Intelligence Ransomware Meta-Core. The cloud environment performs advanced cross-endpoint correlation, sandbox detonation, memory extraction, and real-time key retrieval from active ransomware samples.

Through this distributed learning mechanism, each newly observed threat contributes to the global model used by all participating endpoints. Updated detection rules,

model weight adjustments, and new behavioral signatures are then disseminated to all agents, ensuring that the system continuously adapts to emerging ransomware strains. Finally, the fully integrated system is evaluated through repeated experiments involving real-world ransomware, benign high-I/O applications, legitimate encryption tools, and complex malware exhibiting stealthy or fileless behavior.

Metrics such as detection rate, false-positive frequency, model latency, behavioral correlation accuracy, and ransomware-stage-specific detection time are measured in both offline and live-execution scenarios. By correlating outcomes across multiple monitoring layers (static inspection, DRNN predictions, behavioral graph analysis, honeypot triggers, and cloud validation) the methodology ensures that the final system is rigorously assessed for reliability, adaptability, and operational readiness.

Through this unified methodological approach, the proposed Machine Learning-Based EDR Model establishes a cohesive detection and defense pipeline capable of identifying ransomware and malware at multiple stages of their execution lifecycle, dynamically adapting to environmental changes, and maintaining low false-positive rates through coordinated machine learning inference and cloud intelligence cooperation.

3 Threat Model

The threat model guiding the design and evaluation of the proposed Machine Learning-Based EDR system is structured around the operational realities of modern ransomware and malware campaigns, incorporating attacker capabilities, system assumptions, and the boundaries within which the defense mechanisms operate.

This model defines the classes of adversaries considered, the technical surfaces they can access, the tactics and techniques expected to be employed, and the defensive guarantees the system aims to provide under realistic constraints. The system assumes a motivated adversary with access to a broad spectrum of offensive capabilities, including automated ransomware kits, commodity malware, advanced loader technologies, and polymorphic engines capable of rapidly generating large volumes of functionally identical but syntactically distinct samples. Such adversaries may leverage zero-day vulnerabilities, privilege-escalation routines, or stealthy infection vectors such as email phishing, supply-chain compromise, drive-by downloads, or lateral movement after initial foothold.

The attacker is further assumed to possess the skill to evade signature-based mechanisms by employing dynamic payload generation, runtime packing, shellcode injection, reflective DLL loading, or fileless attack methodologies that exploit legitimate operating system tools (e.g., PowerShell, WMI, mshta) to initiate malicious behaviors. In addition to technical sophistication, it is assumed that the adversary has the logistical capacity to operate command-and-control servers supporting encrypted communication channels, domain fronting, fast-flux networks, or TOR-based isolation.

The primary asset targeted by the adversary is the confidentiality, availability, and integrity of the victim's data and computing resources.

Ransomware adversaries aim to encrypt, exfiltrate, or destroy data, while broader malware operators may seek persistence mechanisms, credential theft, lateral movement, or controlled command execution within the enterprise environment. The system considers endpoints, local data repositories, user-profile directories, shared network locations, and memory-resident data structures as critical resources requiring real-time protection. Additionally, the machine learning pipeline, EDR communication channels, and cloud-based intelligence infrastructure are treated as potential attack surfaces, requiring clear assumptions about trust and resilience.

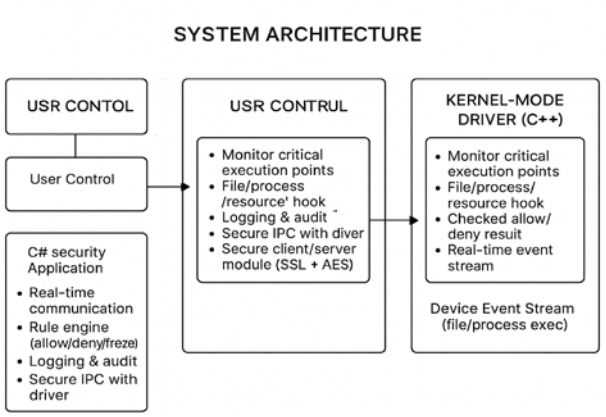

Fig. 2. Architecture of the Machine Learning-Based EDR model showing telemetry collection from kernel and user sensors, local ML-based detection using feature extraction and a Residual Neural Network, and cloud intelligence components that provide behavioral analytics, threat intelligence, and continuous model updates.

The Attacker is assumed to be capable of observing the environment and adapting to defensive actions through iterative reconnaissance. This includes evaluating the presence of monitoring hooks, altering behavioral patterns to delay detection, throttling encryption rates to mimic benign I/O activity, or disabling system protections prior to launching the primary payload.

The threat model therefore includes time-delayed and low-noise ransomware strains designed to bypass simplistic behavioral thresholds, as well as malware variants engineered specifically to confuse machine learning classifiers by generating adversarial patterns, inserting benign-looking behaviors, or mimicking normal user operations.

Despite these attacker capabilities, several constraints are assumed. First, the adversary cannot compromise the cryptographic integrity of the EDR-cloud communication channel, and cannot impersonate the cloud intelligence service or inject false training data into the global learning network. Second, the attacker does not possess the ability to disable all kernel-level monitoring components before initiating malicious actions, as doing so would require privilege escalation events that are themselves detectable. Third, it is assumed that the attacker cannot predict internal ML model weights or inference thresholds, although they may attempt black-box probing through iterative execution of slightly modified samples.

Based on these assumptions, the threat model defines the following defense objectives:

- The system must detect and neutralize ransomware and malware prior to completion of destructive actions such as large-scale file encryption, credential dumping, persistence installation, or lateral movement.
- Given the presence of polymorphic and fileless techniques, the defense system must rely on behaviorally grounded, adaptive detection mechanisms resilient to syntactic evasion tactics.
- EDR agent must respond effectively even when static signatures are absent and when dynamic behaviors exhibit partial overlap with benign activities.
- Additionally, the machine learning architecture must operate with robustness to adversarial noise, maintaining low false positives while ensuring rapid and reliable identification of high-risk events.
- The model furthermore requires that cloud intelligence enhances local detection capabilities by contributing cross-endpoint threat correlation, real-time key extraction from live ransomware samples, and distributed learning updates.
- The system must remain operational under partial network degradation: local ML-based inferences should still function even when cloud access is temporarily limited.
- Finally, the threat model recognizes that full prevention of all intrusion vectors is unattainable; therefore, the system is designed to minimize the operational impact of successful intrusions by suspending suspicious processes, recovering encryption keys when feasible, and limiting the adversary's ability to achieve persistence or lateral expansion.

Under these conditions, the threat model provides a structured adversarial landscape that ensures the proposed EDR solution is evaluated against realistic attacker behaviors, operational constraints, and evolving malware ecosystems.

This framework supports the development of a resilient and adaptive defense model capable of maintaining high detection fidelity in the presence of continuously changing ransomware and malware threats.

4 System Architecture

The proposed system architecture integrates multi-layer endpoint defense, real-time behavioral telemetry acquisition, machine-learning-driven analytics, and coordinated response mechanisms into a unified EDR framework capable of adaptive protection against malware and ransomware.

The architecture is organized around a decentralized collection layer deployed locally on endpoints and a centralized cloud-based intelligence layer responsible for large-scale analytics, model retraining, and signature-agnostic threat reasoning. This architectural duality ensures both low-latency endpoint response and high-capacity global correlation of threat indicators.

At the lowest level of the architecture, lightweight kernel-mode and user-mode sensors are installed directly on the endpoint. These sensors continuously monitor critical subsystems, including process creation events, filesystem operations, registry modifications, network traffic flows, cryptographic API calls, inter-process communication,

and indicators of high-entropy mass-file modification patterns characteristic of ransomware [7]. Each sensor module employs a strictly controlled communication interface that streams fine-grained telemetry toward the local detection engine. In parallel, a secure audit pipeline captures encrypted snapshots of high-risk events for optional forwarding to the cloud intelligence network.

At the core of the local detection engine lies an adaptive machine learning inference module that processes incoming telemetry in real time. The module integrates a hybrid model composed of a lightweight behavioral classifier, a residual neural network (ResNet)-based sequence model [8] for deep temporal analysis, and a feature aggregation layer that fuses static, dynamic, and response-level metadata. By combining static features such as file header entropy or PE structure anomalies with dynamic indicators extracted from system calls and network behavior, the engine can detect malicious activity even when a ransomware family employs polymorphism or obfuscation.

An internal decision-fusion mechanism resolves the outputs of multiple ML components to provide a final risk scoring, enabling deterministic and probabilistic detection paths to complement each other. The architecture incorporates an autonomous enforcement subsystem that executes an appropriate response according to the resulting risk level. When early-stage ransomware behavior is detected, the system can freeze the offending process using kernel-level thread suspension and temporarily isolate its I/O operations. In higher-severity cases, the subsystem can revoke file-system write permissions, disconnect the network interface, or terminate the malicious process outright.

Additionally, the architecture supports a recovery and continuity layer capable of leveraging retained shadow copies and predictive reconstruction models to recover encrypted data. A secure local event ledger ensures full traceability of each enforcement action. The cloud intelligence layer provides global analysis and large-scale model refinement.

Endpoint sensors periodically transmit anonymized telemetry, threat indicators, and partial behavior chains to a distributed analytics cluster. This layer performs large-volume correlation across multiple endpoints, enabling the identification of emerging threat signatures, command-and-control communication templates, and cross-campaign behavioral similarities.

The cloud layer continuously retrains the machine learning models using federated learning and incremental learning techniques, pushing new inference weights to endpoints through encrypted update channels.

This design guarantees that the defense system can evolve in tandem with adversaries, maintaining resilience against zero-day ransomware variants and novel attack chains.

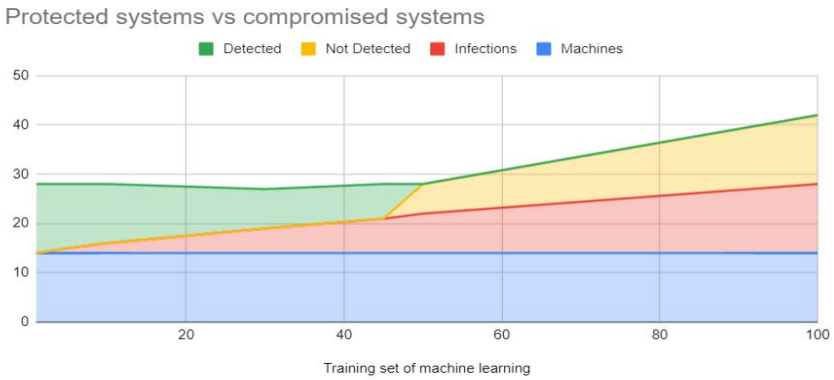

Fig. 3. Benchmark result vs deep residual neural network ransomware detection module.

To ensure system integrity, the architecture employs a cryptographically verifiable update mechanism, certificate-pinned communication channels, and hardware-rooted trust anchors when available. Furthermore, all inter-component communication is authenticated and encrypted, protecting the diagnostic and response channels from adversarial tampering.

The overall architecture thus forms a unified, adaptive, and resilient EDR platform that combines behavioral sensing, deep learning inference, cloud intelligence, and automated response processes to achieve comprehensive defense against modern malware and ransomware threats.

5 Conclusion & Future Work

The present study demonstrates that integrating machine learning with multilayered endpoint telemetry and cloud-driven intelligence could significantly enhance the detection and mitigation capabilities of modern EDR platforms against rapidly evolving malware and ransomware threats. By combining static, dynamic, and behavioral indicators into a unified residual neural network-based analytical model, the proposed approach enables the identification of malicious activity even in environments characterized by polymorphism, obfuscation, and adversarial evasion. The system architecture further enhances resilience by providing real-time process containment, controlled isolation, and predictive data recovery mechanisms, thereby reducing the operational impact of attacks and strengthening endpoint continuity. Experimental evaluations performed across a spectrum of prevalent ransomware families indicate that the model achieves high detection accuracy while maintaining adaptability to emerging attack vectors through continuous cloud-based retraining.

Acknowledgments. The authors of this study are granting a special appreciation for the experimental base and partial funding support to the National Scientific Programme “Security & Defense”. Additional gratitude is also given to the Centre of Competence on Digitisation of the

economy in an environment of Big Data-second stage, established under Grant No. BG16RFPR002-1.014-0013-C01, financed by the Science and Education for Smart Growth Operational Program and co-financed by the European Union through the European Structural and Investment Funds.

References

1. The State of Ransomware 2025, SOPHOS White Paper (2025), http://bit.ly/4avjxli, last accessed 2026/02/11
2. S. Russell, P. Norvig (eds), Artificial Intelligence: A Modern Approach, 4th edn. Pearson (2022)
3. DBIR 2025 Data Breach Investigation Report, Verizon (2025), https://bit.ly/4arMRJl, last accessed 2026/02/11
4. 2025 Ransomware Trends and Proactive Strategies, Veem Insights report, Veem (2025), https://bit.ly/4r8gfvd, last accessed 2026/02/11
5. S. Tafkov, Z. Minchev, Ransomware Detection & Neutralization System, in Proc of X International Scientific Conference Hemus 2020 Research and investment in technology innovation – a crucial factor for defense and security, Defense Institute "Prof. Tsvetan Lazarov", Bulgaria, pp. II-144-II-152, DOI:10.13140/RG.2.2.21029.12009 (2021)
6. D. Medhi, K. Ramasamy, Network Routing (Algorithms, Protocols, and Architectures), 2nd edn. Morgan Kaufmann (2018)
7. A. Orebaugh, G. Ramirez, J. Beale, J. Wright, Wireshark & Ethereal Network Protocol Analyzer Toolkit, 1st edn. Syngress, (2007)
8. yara Documentation Release 4.5.5, Virus Total (2025), https://github.com/VirusTotal/yara/releases, last accessed 2026/02/11
9. M. Neidinger, Python Network Programming Techniques: 50 Real-World Recipes to Automate Infrastructure Networks and Overcome Networking Challenges with Python, Packt Publishing (2021)
10. S. Burns, Hands-On Network Programming with C# and .NET Core: Build Robust Network Applications with C# and .NET Core", Packt Publishing (2019)
11. S. Ludin, J. Garza, Learning HTTP/2: A Practical Guide for Beginners, O'Reilly Media, 1st edn. (2017)
12. W. Odom, CCNA 200-301 Official Cert Guide Library, Cisco Press, 2nd edn. (2024)