2026 Volume: 02 Issue: 06 www.ijrpa.com ISSN 2456-9995 Review Article

International Journal Research Publication Analysis

Page: 01-04

AI-POWERED ZERO-DAY MALWARE DETECTION USING NETWORK TRAFFIC ANALYSIS

***Rutuja Vidyadhar Patil**

India.

Article Received: 22 April 2026
Article Revised: 12 May 2026
Published on: 01 June 2026

***Corresponding Author: Rutuja Vidyadhar Patil**
India.

DOI: https://doi-doi.org/101555/ijrpa.9165

ABSTRACT

As internet usage grows, cyber attacks are becoming more frequent and dangerous. Traditional security systems rely on recognizing known "fingerprints" of viruses, which fails when hackers create brand new "zero-day" attacks. This paper reviews the shift from traditional security to Artificial Intelligence (AI) based systems. We propose a system that uses Machine Learning to learn what "normal" network traffic looks like, flagging anything that behaves strangely. This paper details how the AI model was trained using real network data and outlines a modern, cloud-based design. In this design, lightweight sensors watch the network and send data to a central cloud "brain," keeping computers fast and secure while catching threats in real-time.

KEYWORDS: Network Intrusion Detection, Machine Learning, Random Forest, Zero-Day Malware, Distributed Systems, Packet Sniffing, Cloud Security.

INTRODUCTION

Imagine a security guard at a building. In the past, the guard had a list of known criminals. If someone on the list tried to enter, the guard stopped them. This is how traditional Network Intrusion Detection Systems (NIDS) work. But what if a criminal wears a disguise or is completely new to the city? The guard would let them right in.

In the digital world, these new, unseen threats are called "zero-day malware." To solve this, the cybersecurity field is turning to Artificial Intelligence. Instead of giving the security guard a list of names, we train the guard to recognize suspicious behavior—like someone trying to pick a lock. This paper outlines a project to build an AI-powered security system, detailing how the AI was trained to recognize these behaviors and how it will be deployed across a

network.

Literature Review: How Security is Changing

Over the last few years, researchers have tried many ways to stop network attacks.

The Old Method: Standard firewalls were enough for a long time. However, they are blind to customized or heavily disguised attacks.

The AI Shift: Scientists started feeding massive datasets of network traffic into Machine Learning algorithms. While basic algorithms worked better than old firewalls, they sometimes struggled to process live internet traffic fast enough without making mistakes (false alarms).

The Current Best Approach: Recent studies show that "Ensemble" models—specifically the Random Forest algorithm—perform the best. Random Forest works by creating hundreds of different decision paths and taking a majority vote. It is highly accurate and very good at ignoring confusing background noise on a network.

Teaching the AI: Data and Training Process

Before an AI can protect a network, it must be taught the difference between normal internet browsing and a cyber attack. This process is called Machine Learning.

Where the Data Comes From:

The AI cannot look at standard websites or images; it only understands numbers. We utilized a cybersecurity dataset containing thousands of recorded network sessions. This data includes benign (safe) traffic, like someone watching a video, mixed with malicious traffic, like a hacker trying to guess passwords (Brute Force) or flooding a website to crash it (DDoS). We extracted 15 specific characteristics from this traffic, such as the size of the data packets and the speed at which they were moving.

Preparing the Data (Cleaning):

Real-world data is messy. Sometimes, a network calculation attempts to divide by zero, creating an "infinite" error, or leaves a blank space. Before giving this data to the AI, we cleaned these errors. We also "scaled" the data. Since some network features involve tiny numbers and others involve massive numbers, scaling shrinks everything down to a similar range so the AI does not get confused.

The Training Phase:

We split our cleaned data into two piles. About 70% to 80% of the data was used as a "study guide." We fed this to our Random Forest algorithm, allowing it to find the hidden mathematical patterns of a cyber attack. The remaining 20% to 30% of the data was hidden away to be used later as a "final exam" to test the AI.

The Training Phase:

We split our cleaned data into two piles. About 70% to 80% of the data was used as a "study guide." We fed this to our Random Forest algorithm, allowing it to find the hidden mathematical patterns of a cyber attack. The remaining 20% to 30% of the data was hidden away to be used later as a "final exam" to test the AI.

Initial AI Performance and Results

After the Random Forest model finished studying, we tested it using the hidden data it had never seen before. The model performed exceptionally well.

High Accuracy:

The model achieved a training accuracy of over 99%, meaning it perfectly learned the study material. More importantly, it achieved a **testing accuracy of over 99%** on the hidden data. This proves the model actually learned the concepts and did not just memorize the answers

Precision and Recall:

In cybersecurity, accuracy isn't everything. We also looked at the Classification Report. The model showed incredibly high **Precision** (meaning when it raised an alarm, it was almost always a real attack, avoiding annoying false alarms). It also showed high **Recall** (meaning it successfully caught almost all the attacks without letting them slip through).These metrics prove that the Random Forest algorithm is highly capable of acting as the brain for our security system.

Proposed System Design: Putting the AI to Work

While the AI is smart, it requires computing power. If we install this heavy AI directly on an office laptop, it will slow the computer down. To solve this, we propose a "Distributed Cloud Architecture" split into two parts:

The Local Sensor (The Eyes): We install a very tiny, lightweight script on the user's computer. It has no AI inside it. Its only job is to count the network traffic and send those simple numbers to the cloud.

The Cloud Manager (The Brain): The actual AI model lives on a central cloud server. This server receives the traffic numbers from the laptop. Because the server has plenty of power, it instantly analyzes the numbers. If the traffic matches the patterns of an attack, the cloud dashboard immediately flashes a red alert showing exactly which laptop is targeted.

CONCLUSION

Relying on old lists of known viruses is no longer enough to protect networks. By combining Machine Learning—specifically the Random Forest algorithm—with a modern cloud-based design, we can create a security system that learns and adapts. Our initial training proves that the AI can accurately identify cyber attacks with over 99% accuracy. The next phase of this project will focus on connecting the live local
sensors to the cloud dashboard to prove that this system can catch unknown threats in real-time while keeping the actual computers fast and secure.

REFERENCE

1. Hindy, H., Brosset, D., Bures, M., et al. (2021). "Machine Learning for Zero-Day Malware Detection: A Survey on Challenges and Future Directions." *IEEE Access*, vol. 9, pp. 43452-43472.
2. Gao, Y., Wu, H., Song, B., et al. (2019). "A Distributed Network Intrusion Detection System for DDoS Detection Based on Big Data Framework." *IEEE Access*, vol. 7, pp. 154560-154571.
3. Zhang, Y., & Chen, X. (2024). "Research on Intrusion Detection Based on an Enhanced Random Forest Algorithm." *Applied Sciences (MDPI)*, 14(2), 714.
4. Ahmad, Z., Shahid, A. S., Thulasiraman, P., et al. (2021). "Random Forest Classifier Based Network Intrusion Detection System." *International Conference on Computer and Communication Systems (ICCCS)*.
5. Kumar, V., & Sinha, D. (2021). "A Survey of Cloud Computing Detection Techniques against DDoS Attacks." *Journal of Information Security*, 12(1).