SMARTSHIELD: AI-Driven Cybersecurity Incident Response Automation

Fatma Zahra Gaida, *Member, IEEE,* Molka Weslati, *Member, IEEE,* Wassim Arfa, *Member, IEEE,* Jinen Ben Said, *Member, IEEE,* and Moez Soltani, *Member, IEEE,*

***Abstract*—As organizations face increasingly sophisticated cyber threats, there is a critical need for advanced cybersecurity solutions that leverage Artificial Intelligence to automate incident detection, response, and mitigation. This paper presents an AI-powered infrastructure designed to enhance organizational resilience against evolving cyber threats.**

***Index Terms*—cybersecurity, artificial intelligence , Automation , Incident , threads , response .**

I. Introduction

THE increasing frequency and complexity of cyber-attacks have made network security a critical concern for organizations across the globe. Traditional defense mechanisms, such as firewalls and intrusion detection systems, although effective, are often insufficient in the face of evolving threats. As the sophistication of attacks continues to grow, there is a need for more advanced, automated, and intelligent security solutions that can detect and mitigate threats in real-time, without relying solely on manual interventions. This paper proposes an AI-based (Artificial intelligence based) security infrastructure designed to address these challenges through a multi-layered defense architecture. The solution integrates various security technologies that work in unison to monitor, detect, and respond to security incidents. At its core, the system employs an AI-driven detection mechanism that continuously monitors network traffic and system behaviors. By utilizing machine learning models, the system can identify anomalies and potential threats, providing a proactive defense against attacks. In addition to threat detection, the system incorporates real-time threat intelligence sharing to stay up-to-date with emerging vulnerabilities and attack patterns. Once a threat is detected, the system leverages automated incident response mechanisms, allowing it to mitigate risks swiftly and efficiently. This approach reduces the dependency on human intervention and enhances the system's ability to adapt to new threats dynamically. The proposed system is accessible via a user-friendly web interface, providing network administrators with real-time visualizations of the security status, incident alerts, and recommendations for further actions. This combination of AI, automation, and visualization enables a comprehensive, adaptive, and efficient network security solution.

In recent years, cybersecurity threats have become more frequent, diverse, and sophisticated. Traditional methods, such as signature-based detection and static firewalls, often fail to detect emerging threats like zero-day attacks and advanced malware. They also struggle to handle the massive volume of data generated by modern networks, resulting in delayed threat identification and response. As cyberattacks, such as ransomware and IoT-targeted threats, increase in speed and impact, the demand for real-time responses has risen significantly. To address these challenges, integrating artificial intelligence and threat intelligence into network security is crucial. AI-powered solutions, like machine learning-based intrusion detection systems, can autonomously learn and adapt to new attack patterns, enabling faster, automated threat detection and response. This AI-driven approach enhances threat visibility, shortens response times, and reduces the burden on cybersecurity teams, ultimately improving resilience against sophisticated cyber threats.

The contributions of this paper are as follows: the design and development of an AI-based security infrastructure that integrates advanced detection and automated response capabilities and the use of machine learning for real-time anomaly detection and decision-making.

November 10, 2024

II. PROPOSED SOLUTION

The proposed solution is a comprehensive AI-based security infrastructure designed to enhance threat detection, automate responses, and facilitate easy monitoring through a web interface. The solution architecture leverages microservices to integrate various components into a seamless, end-to-end security ecosystem. This architecture comprises the following primary components:

A. Security Stack

1) Rising Demand for Real-time Incident Response: The network is secured using a dual-firewall configuration, with pfSense positioned between the internet and internal network as the primary firewall. A load-balancer mode is implemented between pfSense and OPNsense to distribute network traffic and ensure redundancy.

2) IDS/IPS (Intrusion Detection and Prevention): The OPNsense firewall serves as an IDS/IPS layer, detecting potential intrusions and preventing malicious access to the network.

F .Z .gaida is a second year computer engineering student
M. Weslati is a second year electrical engineering student
W. arfa is a phd student electrical engineering
J. ben said is an electrical engineer
M. soltani is a professor

3) SIEM (Security Information and Event Management): Wazuh, combined with Snort, functions as a SIEM system, aggregating and analyzing logs to identify potential security events. This SIEM system communicates directly with the MISP threat intelligence platform for real-time threat intelligence, aiding in identifying potential external threats based on known IOCs (indicators of compromise).

B. Anomaly Detection Model

A machine learning model for anomaly detection is employed to identify unusual patterns in network traffic and system behavior that may indicate a threat. This model is trained on historical data to improve its accuracy and reduce false positives. The integration of AI enables the system to detect novel attacks without relying on predefined signatures, addressing a critical limitation of traditional IDS.

C. Web Interface for Monitoring

1) Visualization and Reporting: A web-based interface provides a user-friendly platform for administrators to monitor security events and view the results of the AI detection model in real time. The interface displays alerts, detected anomalies, and incident reports, allowing for effective and centralized security management.

2) Interactive Dashboard: Through an interactive dashboard, users can track incident details, threat levels, and response actions, facilitating a quick overview of the security posture.

D. Incident Response Mechanism

1) Automated Response with D R L : The incident response component leverages a deep reinforcement learning model, specifically a Deep Q-Network (DQN), to automate threat response actions. The reinforcement learning model interacts with the SIEM to receive data on incidents and then takes automated actions, such as reconfiguring firewall rules or applying defensive measures based on the threat severity. This dynamic response mechanism reduces response time, minimizes manual intervention and continuously learns to optimize response strategies

2) Firewall Configuration and Threat Mitigation: The DQN model updates firewall settings in real-time to block malicious IPs, enhance rule settings, and reinforce defensive barriers. This proactive defense mechanism enhances security by dynamically adjusting to evolving threats.

III. Objective

A. Automate Incident Detection and Response

The goal is to automate incident detection and response to enhance the speed and accuracy of threat mitigation. By using AI models and reinforcement learning, the system can quickly identify potential security breaches and respond autonomously, reducing the need for human intervention. This automation improves efficiency, limits human error, and accelerates incident response, providing faster recovery from attacks.

B. Provide Real-time Monitoring and Visual Feedback of Security Events

The system offers real-time monitoring through an intuitive web interface that displays security events and threat data. This visual feedback helps security administrators make quick decisions by presenting key information in easily understandable formats. Immediate visibility into security activities enables faster response times and proactive threat management, ensuring that potential risks are addressed as they arise.

C. Use AI for Advanced Threat Detection and Leverage Threat Intelligence Feeds from MISP

AI is employed for anomaly detection, identifying threats beyond the scope of traditional methods. Machine learning models analyze network patterns to detect previously unknown attacks, while integrating MISP feeds adds a layer of real-time threat intelligence. This combined approach enhances threat detection, ensuring that both emerging and known security threats are addressed more effectively.

D. Develop a Comprehensive Solution that Integrates Firewalls, SIEM Tools and AI into a Single Framework

The solution integrates pfSense and OPNsense firewalls, Wazuh, Snort, and MISP into a unified security framework. This integration ensures seamless communication across all components, optimizing threat detection and incident response. By combining firewalls, SIEM tools, and AI, the system provides a comprehensive, efficient, and adaptive defense against cyber threats.

IV. methodology

A. Related Work

The integration of artificial intelligence (AI) with network security has been a focal point of research in recent years, especially for intrusion detection systems (IDS). Many studies have focused on improving traditional IDS by incorporating machine learning and AI techniques to enhance threat detection accuracy and response times. The use of AI has been particularly effective in identifying previously unknown or novel attack patterns that signature-based methods fail to detect. Existing research has demonstrated the benefits of combining firewalls, IDS/IPS, and threat intelligence platforms to create a more comprehensive and automated security infrastructure. However, challenges remain in reducing false positives, improving detection accuracy and automating the incident response process in real-time. Our solution builds upon this body of work by incorporating deep reinforcement learning (DRL) for dynamic incident response and leveraging threat intelligence feeds for enhanced detection capabilities.

B. Infrastructure Design

The proposed system architecture involves integrating multiple security components to provide a robust, AI-enhanced network defense mechanism. At the core of the architecture is a dual-firewall setup with pfSense acting as the perimeter firewall and OPNsense serving as the IDS/IPS system, configured in a load-balancing mode to ensure high availability and load distribution. Both firewalls monitor incoming and outgoing network traffic, identifying potential threats based on predefined rules and anomalies. These systems communicate with Wazuh, a comprehensive SIEM tool, which aggregates logs from the firewalls and other network devices for centralized analysis. Snort, a well-known IDS/IPS tool, complements this setup by providing real-time packet-level inspection to detect malicious activity. The integration of MISP, a platform for sharing threat intelligence, enhances the system's ability to recognize known attack patterns and respond to emerging threats. An AI-based detection model is incorporated into the system to identify anomalies in network traffic, leveraging machine learning techniques to learn patterns from historical attack data. This model continuously monitors network activity, flagging potential threats based on real-time data. The integration of these components ensures a seamless flow of information across the security stack, facilitating early detection and automated response to threats.

C. AI Model Development

The AI model development process begins with the use of the SNL-KDD dataset, which has become a benchmark for intrusion detection research. Originally developed by DARPA, the dataset simulates network traffic and includes various types of attacks categorized into four classes. The dataset contains 125,973 training instances and 22,548 testing instances, with each row representing an attempted attack with specific attributes such as source IP, destination IP, protocol type, and various traffic metrics. The last column indicates the attack type, which serves as the class label for supervised learning tasks. This dataset is crucial for training and evaluating machine learning models designed to detect and classify attacks based on network behavior. For the threat detection model, we use sequential models based on deep learning architectures. These models are well-suited for capturing patterns in time-series data such as network traffic which is often sequential in nature. The sequential model learns the temporal dependencies between network events, enabling it to detect attack patterns more effectively. In addition to the sequential model, we integrate a Deep Q-Network (DQN), a type of deep reinforcement learning model to automate the incident response process. DQN is particularly well-suited for decision-making tasks where the system learns to take actions (such as reconfiguring firewalls or blocking suspicious traffic) based on the current state of the network. The model is trained using the rewards from successful threat mitigations gradually improving its ability to respond to attacks in real-time. The reinforcement learning model is continuously updated, enabling the system to adapt to new threats as they emerge. This combination of AI models provides both advanced detection capabilities and an automated, adaptive response mechanism, significantly enhancing the security infrastructure's ability to defend against evolving cyber threats.

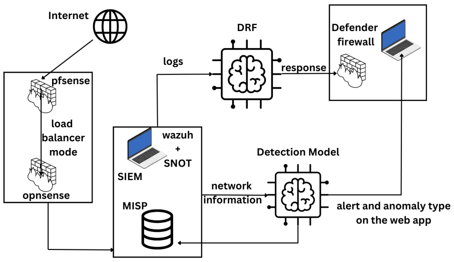

Fig. 1. Solution architecture

V. Implementation

A. Solution Components

The implementation of the proposed system is based on a microservices architecture, where each component functions as an independent, loosely coupled service communicating through well-defined APIs. This architecture ensures scalability, flexibility, and ease of maintenance while providing robust security across the entire infrastructure.

1) Firewall Integration: The firewall integration of pfSense and OPNsense ensures robust protection of the network by acting as the first line of defense against external threats. pfSense serves as the gateway firewall, filtering incoming and outgoing traffic based on security policies, while OPNsense is deployed in load-balancer mode, providing additional IDS/IPS capabilities for real-time detection and prevention of intrusions. These firewalls are designed to communicate with each other to ensure synchronized security across the entire infrastructure, creating a layered defense model. The firewalls are configured to detect and block unauthorized access, malicious traffic, and other potential threats before they can infiltrate the internal network.

2) Security Information and Event Management (SIEM): The integration of Wazuh and Snort as part of the SIEM system plays a pivotal role in collecting, analyzing, and monitoring security events. The Wazuh agent collects log data from firewalls and other network devices, correlating and analyzing it to detect potential security threats. Snort, a widely used intrusion detection system (IDS), is integrated to monitor network traffic for suspicious activity and generate alerts for any potential intrusions. Both tools are linked with MISP (Malware Information Sharing Platform), which aggregates threat intelligence feeds that provide context and real-time updates on new vulnerabilities, helping to improve detection accuracy and speed. This SIEM setup ensures that potential security incidents are promptly identified, and appropriate actions can be taken to mitigate risks.

3) AI-Based Detection: The AI-based detection system leverages machine learning algorithms to analyze network traffic and detect potential threats through anomaly detection. Using the SNL-KDD dataset, which contains labeled instances of both normal and attack traffic, the AI model is trained to identify deviations from typical traffic patterns that may indicate a security threat. Once trained the model can process real-time data from network traffic to spot suspicious activities such as denial-of-service (DoS) attacks, intrusions, and malware infections. By using machine learning the system can continuously improve its detection capabilities , providing a proactive defense against emerging threats while minimizing false positives.

TABLE I
MODEL TRAINING AND VALIDATION RESULTS

Epoch	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
1	0.856945	0.945661	0.656642	0.208663
2	0.959731	0.961464	0.150374	0.162043
3	0.967147	0.964513	0.110510	0.129581
4	0.973385	0.967286	0.088226	0.122090
5	0.975495	0.972785	0.077256	0.123073
6	0.977821	0.974144	0.069810	0.118788
7	0.979803	0.970952	0.065580	0.116089
8	0.981356	0.971167	0.055580	0.115074
9	0.980089	0.973662	0.055380	0.190729
10	0.982534	0.976013	0.048660	0.108631
11	0.983504	0.975049	0.046528	0.115916
12	0.985342	0.973904	0.047007	0.121198
13	0.983989	0.972630	0.040482	0.123656
14	0.985168	0.976157	0.039264	0.126248
15	0.985821	0.976250	0.038204	0.122549
16	0.985168	0.975880	0.037691	0.132518
17	0.985861	0.972757	0.034551	0.130327
18	0.985792	0.972553	0.034638	0.127881
19	0.987316	0.975049	0.032717	0.122779
20	0.987316	0.975049	0.032717	0.136208

4) Incident Response and Reinforcement Learning: The Deep Reinforcement Learning (DRL) model, specifically a Deep Q-Network which has been coined to refer to this version of the Q-learning algorithm. DQN is the main model used in our experimentation below. DQN models apply an experience replay approach as shown to improve the prediction power of the model :

Algorithm 1 Q-Learning Algorithm

```
Input: a policy that uses the action-value function, $\pi(a|s, Q(s,a))$
Initialize $Q(s,a) \leftarrow 0$, for all $s \in \mathcal{S}$, $a \in \mathcal{A}(s)$
for each episode do
    Initialize environment to provide $s$
    repeat
        Choose $a$ from $s$ using $\pi$, breaking ties randomly
        Take action $a$, and observe $r$, $s'$
        $Q(s,a) \leftarrow Q(s,a) + \alpha\left[r + \gamma \max_{a' \in \mathcal{A}(s')} Q(s',a') - Q(s,a)\right]$
        $s \leftarrow s'$
    until $s$ is terminal
end for
```

DQN, automates the incident response process by analyzing detected threats and taking action to mitigate them. Once the AI model flags a potential attack, the DRL model determines the best course of action, such as reconfiguring firewall rules or blocking malicious IP addresses, to prevent further damage. The system continuously learns from past responses, allowing it to adapt and improve over time. By automating this process, the system reduces the need for human intervention, ensuring faster and more consistent responses to security incidents, while also strengthening the overall security posture of the network. The algorithm is based on the Q version of the temporal difference equation :

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left(r + \gamma \max_{a_s \in \mathcal{A}(s')} Q(s', a_s) - Q(s,a) \right) \tag{1}$$

TABLE II
DQN MODEL CONFIGURATION

Parameter	Description/Value
Discount factor (gamma)	0.73
Learning rate (alpha)	0.5
Epsilon greedy policy factor	0.5
Experience replay	A three-layer MLP deep learning model
Actions	‘fix’ ,dostherapy(), probetherapy(), u2rtherapy(), r2ltherapy()
Rewards	Reward for ‘fix’ =0, reward for action executed =1
States	The states are represented by the input stream vectors. The stream moves from one state to another based on the value of passing vectors

B. User Interface and Visualization

The web interface serves as the main control and monitoring dashboard for administrators. Built on a microservices architecture, the interface communicates with the security stack to display:

1) AI Detection Results: The real-time output of the AI detection model, showing the presence of any threats and their classifications (DoS, intrusion, malware ...).

2) Security Events: A comprehensive log of all security events, including alerts from Wazuh, Snort, and the AI model, providing administrators with a holistic view of network health and activity.

3) Incident Response Actions: Automated responses triggered by the reinforcement learning model, including changes to firewall configurations, blocking of malicious traffic, and other preventive measures.

This microservices approach allows the system to scale efficiently while enabling real-time processing and decision-making. Each component is independently scalable and can be updated or replaced without affecting the entire system, making it both adaptable and resilient to evolving cyber-security threats. By leveraging microservices, the system is designed for continuous operation and rapid response to security incidents, significantly reducing human intervention while maintaining high security standards.

VI. CONCLUSION

This paper presents an AI-driven cybersecurity incident response automation system that integrates advanced threat detection, real-time monitoring, and dynamic automated response mechanisms. By leveraging machine learning and deep reinforcement learning , the proposed solution enhances the speed and accuracy of incident detection and mitigation, offering a scalable, adaptable defense framework for modern organizations facing increasingly sophisticated cyber threats. The integration of a multi-layered security architecture, combining IDS/IPS , SIEM, and threat intelligence platforms, ensures comprehensive coverage against a wide range of attack vectors. The system's ability to automate incident response reduces reliance on human intervention, minimizes response times, and improves overall security posture, making it an essential tool for organizations looking to enhance resilience against evolving cyber threats. The incorporation of a user-friendly web interface for real-time monitoring and reporting ensures that security teams can efficiently oversee the system's performance, fostering faster decision-making and more effective threat management. This approach provides a proactive, scalable solution that not only addresses the limitations of traditional security systems but also ensures preparedness for emerging threats.

Acknowledgment

The authors would like to express their sincere appreciation to everyone who contributed to the success of this work. They are also grateful to our colleagues and mentors for their continuous support, valuable feedback, and guidance throughout the course of this study. Their insights were essential in refining our approach and ensuring the quality of the work presented.

References

[1] Adversarial Challenges in Network Intrusion Detection Systems: Research Insights and Future Prospects.Available online: https://ar5iv.org/abs/2409.18736

[2] . Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. Available online: https://www.mdpi.com/2076-3417/9/20/4396.

[3] Current trends in AI and ML for cybersecurity: A state-of-the-art survey. Available online: Full article: Current trends in AI and ML for cybersecurity: A state-of-the-art survey.

[4] Optimizing cybersecurity incident response decisions using deep reinforcement learning. *IJECE*, vol. 12, no. 6, pp. 6768–6776, Dec. 2022. ISSN: 2088-8708, DOI: 10.11591/ijece.v12i6.pp6768-6776.

[5] A Deeper Dive into the NSL-KDD Data Set (A Deeper Dive into the NSL-KDD Data Set — by Gerry Saporito — Towards Data Science ref) ://towardsdatascience.com/a-deeper-dive-into-the-nsl-kdd-data-set-15c753364657.

Fatma Zahra Gaida a second year computer engineering student and ieee cs ensit sbc vice chair

Molka Weslati a second year electrical engineering student and ieee cs ensit sbc treasure

Wassim Arfa is a PhD student at University of (ENICar). He is a member of the research laboratory: RIFTSI at the ENSIT University of Tunis. His research interests include robotics, machine learning, and deep learning.

Jinen Ben Said An Electrical Engineer , joined IEEE in 2021 and quickly took on leadership roles, serving as Secretary from 2022 to 2023, and later as Chair and Treasurer for IEEE AESS ENSIT SBC. She is currently the Chair of the AESS Student Branch Chapter for 2023-2024.

Moez Soltani is an associate professor at(ENSI) and is the IEEE CS ENSIT SBC Advisor.