Application of Diverse Techniques for Zero-Day Management: A Review Approach

[*]Nyia Okechukwu Cosmos., Gilbert Aimufua

Centre for Cyberspace, Nasarawa State University, Keffi (NSUK), Nigeria

***Corresponding Author**

DOI: https://doi.org/10.51584/IJRIAS.2025.1005000126

Received: 18 May 2025; Accepted: 22 May 2025; Published: 25 June 2025

ABSTRACT

The increasing sophistication and frequency of zero-day attacks pose significant challenges to traditional cybersecurity defences, necessitating innovative approaches for timely detection and mitigation. This study presents a comprehensive systematic literature review on the application of deception security techniques, machine learning algorithms to enhance zero-day attack management. By analysing recent advances in deception-based intrusion control techniques, honeypot and adaptive camouflage, alongside the application of machine learning models in improving threat detection accuracy and response adaptability. Key observations reveal that deception techniques offer early threat detection, behaviour-based analysis, reduced false positives, and effective attack diversion. These attributes make deception a scalable and proactive approach to enhancing cybersecurity defences against zero-day threats. Data of UGRansome was collected which contain classes of zero-day, and then Random forest was selected and trained to generate behavioural analytical model. The model performance recorded 99% accuracy and was validated through comparism with other literatures. The zero-day attack detection model generated was recommended as decision based for deception model for improved zero-day management.

Keywords: Zero-Day Attacks; Deception Security; Machine Learning; Intrusion Detection; Cybersecurity

INTRODUCTION

Globally, one of the main challenges facing computer system networks is Zero-Day Attack (ZDA). Sarhan et al. (2022) defines ZDA as an unknown cybersecurity vulnerability, which hackers exploit to illegally penetrate and attack a network. The term “Zero-day” implies that these vulnerabilities are hidden or not known to the network administrator, hence leaving them zero time of notice to fix the flaws before they are exploited for attack. As the sophistication of cyber-attacks has continued to evolve, ZDA has become increasingly challenging to detect, mitigate and manage (Ferguson-Walter, 2019). Traditional solution depends on antivirus, firewall, and intrusion detection systems; however, these techniques are not reliable for the management of attack (Teymourloueiet al., 2023).

Among these three components of zero-day, vulnerability is the most prioritized for cyber criminals. According to Kaspersky (2022) and Peppes et al. (2023), zero-day vulnerabilities are highly valuable in the underground market when auctioned by threat actors (groups of hackers who actively search for vulnerabilities in computer networks, software or hardware for exploitation), and are exploited for ZDA. For instance, in 2022, the Google project zero team reported 18 different ZDA (Google, 2022). Another example is the Microsoft CVE-2016-067 vulnerability threat on windows machine (Microsoft, 2022).Most recently, deception technique has dominated techniques for the management of ZDA, but suffers issues of false positive, poor management of threats, and inability for large scale network deployment, thus leaving critical weakness in the current security solutions (Peppes et al., 2023; Tian and Zhao, 2024; Kumar and Subbiah, 2022).

Deception technology is defense tactics which employ deceptive tools to divert attacker away from original network infrastructure to a decoy facility, and have been engages for ZDA detection, monitoring and mitigation

(Adel et al., 2019; Shalaginov et al., 2016; Oluoha et al., 2021; Bowen et al., 2018). In Mohan et al. (2022), deception technique was identified as one of the most researched defense strategies in cyber security studies due to several advantages it provides, particularly decoy and threat intelligence.

Popular deception methods include Honey-X, camouflaging, mimicking, etc. (Oluoha et al., 2021). Among the deception methods, Honey-X (Abe et al., 2021) has been widely used by researchers (Happa et al., 2021; Chiang et al., 2018; Xingshen, 2023) for ZDA management. Honey-X involves setting up a decoy of the network infrastructure using techniques such as honeypot, honeytoken, honey-web, and honey-net, to trap the attacker into the decoy facility. Honey-X overwhelms the attackers and wastes their resources, by creating ambiguity for the adversaries and hampering them from achieving their criminal goal; however, these techniques suffer many limitations such as complexity in scaling across large network infrastructure, lack of adaptive intelligence, false positive behaviour and cannot manage vulnerability. There is a pressing need for an advance deception technology for zero-day attack management, and this study will review the application of advance intelligent approaches capable of handling ZDA and deference to reinforce security of Information Technology(IT) systems.

LITERATURE REVIEWS

This section discussed the literature review of literatures in three broach categories, starting with the review of relevant literatures on deception security for zero-dayattack management considering the general deception strategies such as honeypot, mimicking, camouflaging, etc. then the review in another section discussed the application of Machine learning (ML) algorithm for deception security in managing zero-day vulnerabilities was discussed in the third section.

Review of Relevant Literatures on Zero Day Detection

SakthiMurugan et al., (2024) presents a study on the use of machine learning approach for the assessment of zero-day vulnerability. The purpose of this research is to offer a dependable method for identifying zero-day vulnerabilities and invaders in software systems. To identify anomalous data patterns, the proposed approach combines an auto-encoder model with a Deep Learning (DL) model. Furthermore, a model for outlier detection will be created that compares the single class-based Support Vector Machine (SVM) method to the autoencoder model. Two popular IDS datasets are used to assess the suggested model. The CICIDS2017 dataset, created by CIC in Canada, includes a wide spectrum of recent instances of attacks, both inside and outside of buildings. There are 122 neurons in total between the ANN's input and output layers.Additionally, there are three hidden layers that comprise the autoencoder-optimized design of the NSL-KDD dataset, with 100, 60, and 100 neurons in each. A 1024 batch size is recommended. Other ideal parameters include fifty epochs, 0.001 regularisation of L2, and the average absolute error value of loss.

Walter et. al., (2021) aimed to ascertain the effects of deception on a cyber attacker that was trained using reinforcement learning (RL) to achieve its goal within the autonomous cyber defence environment. The study used CyberBattleSim which was originally devised to research autonomous agents using RL in computer network but did not include any deception concepts which brought about the integration of decoys, honeypots, and honeytokens into the codebase of the CyberBattleSim. Reward function, attacker wins, wasted resources, and defender detections were used as metrics for evaluation in the study. Results presented in the study showed that modelling cyber deception into CyberBattleSim can be effective on analysing attacker behaviour. However, there should also be autonomous defender agents that can take deceptive responses and also more sophisticated attackers so that the experiment can be applicable to real-world attack behaviour.

Hindy et al., (2020) presents a study on the use of deep learning for the detection of zero-day attack effectively. An autoencoder implementation for identifying zero-day attacks is suggested in this study. The goal is to create an IDS model with a high recall and a minimally acceptable miss rate (false-negatives). CICIDS2017 and NSL-KDD, two popular IDS datasets, are utilised for evaluation. The work benchmarkedthe model's performance against a One-Class Support Vector Machine (SVM) to show how effective it is. The study demonstrates how well a One-Class SVM performs in situations where zero-day attacks may be distinguished from typical behaviour. The encoding-decoding abilities of autoencoders are very advantageous to the

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue V May 2025

suggested paradigm. The outcomes demonstrate that autoencoders are a good fit for identifying intricate zero-day attacks. For the NSL-KDD dataset, the results show an accuracy of 89–99% in zero-day detection, while for the CICIDS2017 dataset, it ranges from 75–98%.

Oluoha et al., (2021) surveyed on cutting edge trends in Intrusion Detection Systems (IDS) based on deception. This article reviews current trends and implementation approaches in deception-based intrusion detection systems and provides a comprehensive overview of the deception technology landscape. It is also looked into how to mitigate the use of deception-based cyber security measures. This book provides a comprehensive introduction to deception-based technology, including topics such as taxonomies, psychological notions of deception, applications of deception, and legal and ethical considerations. By using deceit, a cyber security expert can improve defence capabilities and more accurately detect and attribute cyberattacks by continuously learning about potential attackers at different stages of the cyber-attack chain.It should be highlighted that the use of deception in cyber security defences may potentially introduce actual hazards, which need to be thoroughly evaluated, examined, and taken into account before being implemented.

Kovářová (2024) explored the impacts of zero-day attacks on machine learning (ML) and deep learning (DL) algorithms. This study explores the types, causes, effects, and potential countermeasures of zero-day attacks on ML/DL algorithms, offering a thorough review of the subject. The term and concept of zero-day attacks are presented at the outset of the work to provide readers a clear knowledge of this new threat. After that, a review of previous studies on zero-day attacks on ML/DL algorithms is conducted, with a particular emphasis on three major types of attacks: model stealing, adversarial input, and data poisoning. The possible effects and dangers of these attacks on different application domains are also covered in the study. The article concludes by outlining a few potential defence strategies against zero-day attacks on ML/DL algorithms.These include methods for protecting the privacy of the training data through federated learning, techniques for detecting anomalies in the data or the algorithm's behaviour, methods for validating and verifying the model to ensure the algorithm's correctness and robustness, and methods for adding noise to the data or the algorithm's outputs to prevent information leakage.

Topcu et al., (2023) researched on the use of TensorFlow for the detection of social media zero-day attack. The goal of this project is to quickly detect vulnerabilities and thwart zero-day attacks by analysing data from the Twitter platform and applying machine learning techniques, like word categorization. The processing and conversion of unprocessed Twitter data was handled by TensorFlow, which led to notable efficiency gains. The Natural Language Toolkit (NLTK) tool was also added into the study to extract specific words in a variety of languages. According to the study's findings, it used the technique to detect zero-day attacks with an 80% success rate. Through the use of information that people have provided publicly, pertinent security providers can be notified in a timely manner.

Morozov et al., (2023) presents a study on the use of honeypot and cyber deception approach as a tool for the detection of cyber-attacks on network infrastructure. An overview of the benefits of employing cyber deception and honeypot/honeynet solutions for industrial Internet of things (IoT) and general-purpose networks is provided in this article. In order to gather static data on the attack pathways and tactics used by attackers, the study intends to concentrate more research on the development of a believable IoT honeynet network. This network will include standard setups and settings for IoT networks in Ukraine. The ability to investigate the toolset of attackers for identifying honeypots and honeynets in IoT networks will be made possible by expanding and enhancing the functionality of this network in conjunction with the application of machine learning technologies to produce believable intra-network traffic.

Cen et al., (2024) presents zero-ran sniff which is based on zero-shot learning for an early detection method for zero-day ransomware.In order to detect zero-day ransomware attacks early on, this study presents Zero-Ran Sniff (ZRS), an early zero-day ransomware detection approach based on zero-shot learning. ZRS uses executable files' portable executable header (PE header) functionality to detect malware. It consists of two stages: the Self-Attentive mechanism-based Convolutional Neural Network Inference Stage (SA-CNN-IS) and the Auto-Encoding Network-Based Core Attribute Learning (AE-CAL) stage. Self-encoding networks are used to extract the essential characteristics of both known and unknown ransomware classes during the AE-CAL

stage, and ransomware is identified during the SA-CNN-IS phase. As a consequence of the system's installation, 98.47% recall and 96.31% accuracy were attained.

Singh et al., (2019) researched on the detection and prioritization of zero-day vulnerabilities. This study presents a framework that provides an integrated strategy for detecting zero-day attacks and prioritising them (based on likelihood). The suggested methodology uses a probabilistic method to rate the severity of each discovered zero-day vulnerability and identify the zero-day attack path. It is a hybrid detection-based technique that finds network issues that have not yet been discovered. The work implemented the suggested framework in the network environment of the Vikram University campus in India in order to assess its performance. The experimental results demonstrated a 96% detection rate for zero-day attacks with a 0.3% false positive rate, indicating great promise for the framework.

Reddy et al., (2024) presents a proactive approach base for zero-day vulnerability on Internet of Things (IoT) devices using reinforcement learning. The suggested technique makes use of network metadata and real-time telemetry data from Internet of Things devices to speculate on possible Zero-Day vulnerabilities. through the use of a deep reinforcement learning architecture to train an agent. With the use of a deep reinforcement learning architecture, the model gains the capacity to make hypotheses that result in reliable vulnerability identification. The research then uses a deep reinforcement learning architecture with Proximal Policy Optimisation (PPO) to optimise the agent's policy in order to improve the accuracy of vulnerability detection. Using logical predicates, the action space entails creating conjectures that indicate potential weaknesses. An incentive function is designed to encourage meaningful conjectures. The agent's advancement from basic to more complicated hypotheses is aided by curriculum learning, and ongoing learning guarantees adaptability to shifting IoT network dynamics. The model's capabilities are further enhanced by the incorporation of human skills.The result of the system presents that the system achieved a detection accuracy of 0.925.

Mohamed et al., (2024) comprehensively reviewed on the advances of applying machine learning techniques for detection and mitigation of zero-day vulnerabilities. This research offers a thorough analysis of the most recent machine learning techniques used in the identification and mitigation of zero-day exploits, a serious cybersecurity risk. It analyses the development of machine learning methods from simple statistical models to complex deep learning frameworks and assesses how well they detect and counteract zero-day threats. Along with issues like data scarcity, false positives, and the ongoing arms race against cyber attackers, the integration of ML with other cybersecurity methods to create effective, adaptive defensive systems is also examined. Innovative approaches that improve prediction and real-time response times are given particular consideration.In order to provide researchers, cybersecurity experts, and policymakers with greater tools for their continuous fight against zero-day exploits, this study attempts to summarise current trends and forecast future advancements in machine learning technology. This review provides a roadmap for the ongoing development and improvement of machine learning applications in the battle against zero-day attacks, laying the foundation for future research.

Roumani (2021) presents an empirical analysis of patching zero-day vulnerabilities. Examining the effects of additional, as-yet-undiscovered factors on the patch release time of zero-day vulnerabilities is the primary goal of this study. The work applied survival analysis technique on zero-day vulnerability dataset collected between 2010 and 2020. The influence of vulnerability attack vector, attack complexity, necessary privileges, user involvement, scope, confidentiality, integrity, and availability are all examined in the model in relation to the timeliness of patch releases. Results indicate that if a zero-day vulnerability causes a shift in scope and impacts additional vendors, products, and versions, it has a higher chance of being patched on schedule. Nevertheless, if a zero-day vulnerability affects confidentiality and calls for privileges, it is unlikely to be patched in a timely manner.The sub-analyses also show how different products and vulnerability categories have different patch release dates. One of the study's limitations is that the majority of the sample's product types are closed-source software. As a result, the findings might not be relevant to open-source initiatives. The authors advise conducting survival analyses to assess whether patch release times for open- and closed-source projects vary as more data become available.

Kumar and Subbiah (2022) researched on the use of Shapley ensemble boosting and bagging approach for zero-day malware detection and effective malware analysis. This study reduced misclassification to improve

the performance of bagging and boosting machine learning models. Shapley values of features aid in identifying the most important characteristics for each prediction made by the machine learning model and are an accurate depiction of the contribution of each feature. Shapley values are converted to a probability scale in order to identify the most important attributes for each prediction made by a trained machine learning model and to correlate with its prediction value. Inductive rules can be constructed using the trend of top features obtained from false positive and false negative predictions made by a trained machine learning model.Two gradient boosting decision tree (GBDT) boosting machine learning (ML) models, such as XGBoost and LightGBM, and two bagging ML models, such as Random Forest and Extra-tree forest, are used in this work's trials. As potential zero-day malware in the future, XGBoost consistently produced the top performance results, with accuracy scores of 97.87 and 97.50 from the D2 and D3 datasets, respectively.

RESEARCH METHODOLOGY

To ensure a comprehensive, objective, and replicable examination of the existing body of knowledge, this study adopts the Systematic Literature Review (SLR) methodology which applies a clearly defined process for identifying, evaluating and synthesizing relevant literature in a transparent and structured manner. The application of SLR methodology for this study was guided by the need to critically evaluate the existing deception-based security strategies and machine learning techniques for managing zero-day attacks and finally to examine cyber-attacks that exploit network vulnerabilities from a technical and strategic perspective. The methodology adopted a systematic search process by acquiring studies from digital databases like IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library and Google Scholar considering keywords on “zero-day attacks, machine learning and zero-day vulnerability detection”. The framework for this review illustrates the expected relationship between underlying variables of zero-day attack and it describes the relevant terms for the research process; show the relationships in addressing the research problem and how they come together to draw coherent conclusions as depicted in the framework in Figure 1.

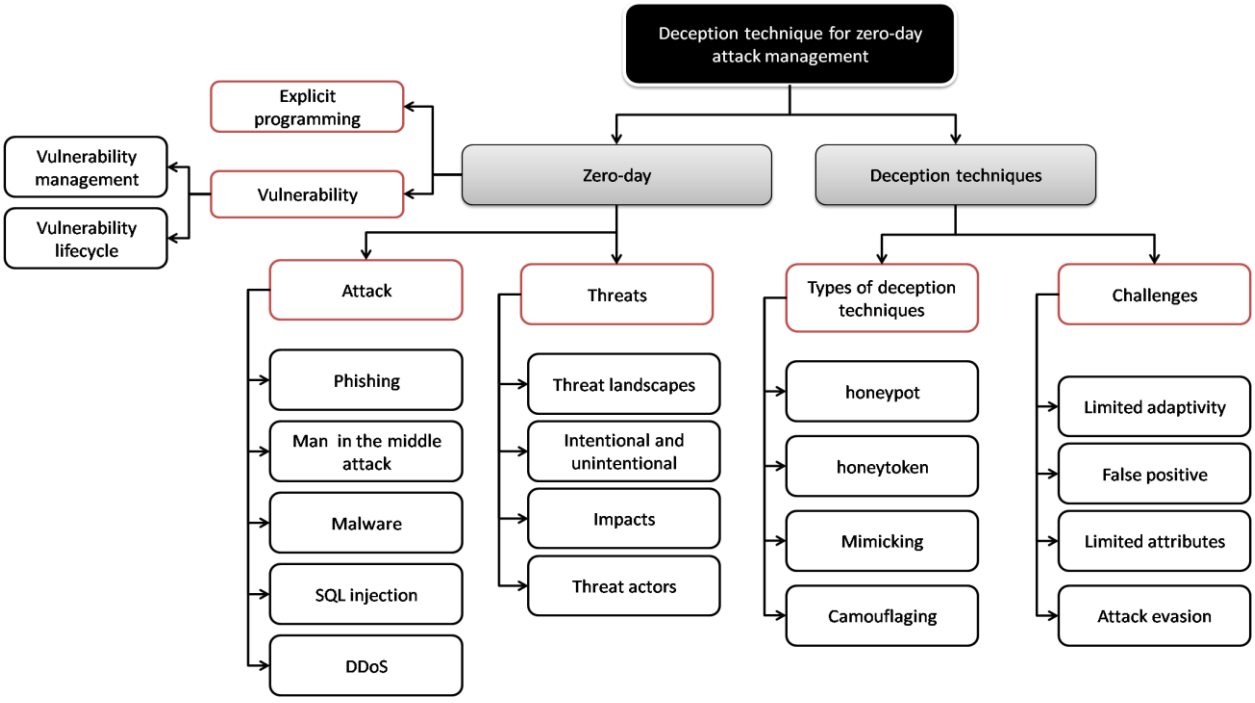

Figure 1: The Research Methodology Framework

The Concept of Zero-Day Attacks

A zero-day attack is a cyber-attack exploiting a vulnerability that has not been publicly disclosed. Due to the challenges associated with zero-day attacks, authors conducted a systematic study to learn the characteristics of zero-day attacks from the data collected from real hosts and identify executable files linked to exploits of known vulnerabilities (Peppes et al., 2023). Zero-day attacks are a type of cyber threat that exploits software vulnerabilities unknown to the public or the software vendor (Sarham et al., 2023). Hackers take advantage of

this knowledge gap to launch attacks that can be exceptionally difficult to detect and defend against. These attacks pose a severe risk to organizations, as traditional security measures (Kumar and Subbiah, 2022) often lack the signatures or patches needed to thwart them, emphasizing the need for proactive threat detection mechanisms. The major classes of zero-day attacks are presented in Figure 2.

Figure 2: Classes of Zero-Day Concepts

Zero-day vulnerabilities

Zero-day vulnerabilities refer to security flaws in software, hardware, or firmware that is exploited by attackers before the affected vendor becomes aware of the issue (Manish, 2023). These vulnerabilities can be particularly concerning because they leave users exposed to potential cyber threats during the period between discovery and the release of a security patch (Ferguson-Walter, 2021)

Vulnerability lifecycles

The life cycle of a zero-day vulnerability unfolds in distinct stages:

i. **Discovery:** The initial phase involves the identification of the vulnerability, typically by a security researcher or, in some cases, a malicious actor. This discovery marks the moment when the vulnerability becomes known, but it is crucial to recognize that during this stage, there is often limited awareness within the broader cybersecurity community.
ii. **Exploitation:** Following the discovery, malicious actors exploit the identified vulnerability to achieve various objectives, such as financial gain, espionage, or other malicious purposes. This stage involves the active and unauthorized use of the vulnerability to compromise systems or data.
iii. **Detection:** Security tools, incident response teams, or vigilant researchers detect the exploitation of the vulnerability. This phase is critical for identifying ongoing attacks and understanding the potential impact on affected systems.
iv. **Disclosure:** Responsible disclosure is a pivotal step where the vulnerability is reported to the vendor or a trusted coordination body. This responsible disclosure aims to provide the vendor with the necessary information to develop a solution while minimizing the risk of widespread exploitation.
v. **Patch Development:** Upon receiving information about the vulnerability, the vendor initiates the development of a patch or security update to address the identified flaw. This process involves creating a solution to mitigate the risk posed by the vulnerability.
vi. **Patch Deployment:** Organizations, upon receiving the patch from the vendor, deploy it to their systems. This step is crucial for closing the security gap and safeguarding against potential exploitation. Timely and widespread deployment of patches is essential to minimize the window of opportunity for attackers to leverage the vulnerability.

Mitigating Vulnerabilities

To address zero-day vulnerabilities effectively, organizations should implement a comprehensive security strategy with the following key elements (Manish, 2023):

a. Embrace Robust Cybersecurity Frameworks: Adhere to robust Cybersecurity Frameworks such as NIST CSF, ISO/IEC 27002, ACSC Essential 8, and CIS, incorporating standards, guidelines, and best practices to mitigate cyber risks and threats effectively.
b. Establish a 24X7 Security Operations Centre (SOC): To enhance detection and response capabilities, set up a Security Operations Centre that operates around the clock, ensuring a proactive approach to identifying and mitigating potential security incidents.
c. Implement Network Segmentation: Isolate critical systems by employing network segmentation, effectively segregating them from less secure areas of the network to limit the potential impact of security breaches.
d. Prioritize Regular Patching: Apply software patches promptly as they become available to address known vulnerabilities and enhance the overall security posture.
e. Deploy Behavioural Analytics Tools: Utilize behaviour-based security tools to detect and respond to anomalous activities, providing an additional layer of defence against evolving threats.

Zero-day attacks

Zero-day attack encompass a broad range of malicious activities as presented in Figure 3 conducted in the digital realm with the intent to compromise, disrupt, or gain unauthorized access to computer systems, networks, or data.

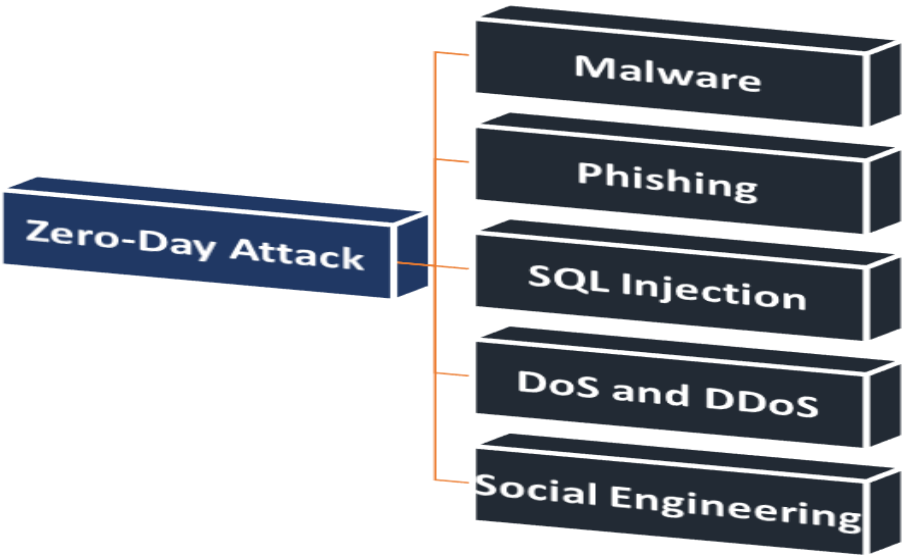

Figure 3: Zero-Day Attack Types

Various types of cyberattacks exist, each employing distinct techniques and targeting different vulnerabilities. Some common cyberattacks include (Guo, 2023):

a. **Malware Attacks:** Malicious software, or malware, is designed to harm or exploit computer systems. This includes viruses, worms, trojan horses, ransomware, and spyware. Malware can be delivered through infected email attachments, malicious websites, or compromised software.
b. **Phishing Attacks:** Phishing is a social engineering technique where attackers use deceptive emails, messages, or websites to trick individuals into providing sensitive information such as usernames, passwords, or financial details. Phishing attacks often disguise themselves as trustworthy entities to manipulate users.
c. **Denial-of-Service (DoS) and Distributed Denial-of-Service (DDoS) Attacks:** DoS attacks overwhelm a system, network, or service with a flood of traffic, causing it to become unavailable. DDoS attacks involve multiple compromised devices, forming a botnet to amplify the volume of malicious traffic, making it harder to mitigate.
d. **SQL Injection:** SQL injection attacks target vulnerabilities in web applications by injecting malicious SQL code into input fields. This can lead to unauthorized access to databases, data manipulation, or even the deletion of sensitive information.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue V May 2025

e. **Social Engineering Attacks:** Social engineering involves manipulating individuals into divulging confidential information or performing actions that may compromise security. This can include pretexting, baiting, or quid pro quo tactics.

Zero-day Threats

Threat refers to any potential danger or harmful event that may exploit vulnerabilities in a system, network, or organization's security. Threats can take various forms, including software vulnerabilities, insecure configurations, or even the presence of malicious actors.

Threat actors

Threat actors refer to the entities or individuals capable of posing a threat, and these actors can be both internal, such as employees, and external, including hackers and malicious software. Understanding the motivations and capabilities of these actors is crucial for tailoring security measures effectively. Internal actors may have privileged access, making them potential threats, while external actors may have diverse motivations ranging from financial gain to cyber-espionage or activism.

Intentional threats

Intentional threats, characterized by deliberate actions with malicious intent, constitute a significant aspect of the threat landscape. Cyberattacks, hacking attempts, malware deployments, and social engineering fall under this category, with threat actors seeking unauthorized access, data theft, or system disruption (Sharukh, 2020). The threat landscape itself is dynamic, encompassing the ever-evolving scope and variety of potential threats. It includes both known and emerging threats, necessitating a proactive and adaptive approach to cybersecurity. Organizations need to stay vigilant, continually assess the evolving threat landscape, and update security measures accordingly (Eze et al., 2022; Sarhan et al., 2023).

Unintentional threats

Unintentional threats, while lacking malicious intent, can still pose significant risks to information security. Accidental data breaches, system mis-configurations, or natural disasters leading to data loss are examples of unintentional threats. Recognizing that not all threats are driven by malicious intent underscores the importance of a holistic security strategy that addresses both intentional and unintentional risks (Eze et al., 2022).

Impacts

The impacts of security breaches or threat events are diverse and extend beyond the immediate compromise of information. Financial losses, reputational damage, legal repercussions, and operational disruptions are among the various categories of impacts that organizations must consider. Understanding the potential consequences allows for a more comprehensive risk assessment and aids in prioritizing security measures based on the potential severity of impacts.

Deception Technique for Cyber Security

Deception techniques in cybersecurity are an evolving and effective approach to bolstering an organization's defences against cyber threats. These techniques involve the art of creating deceptive elements within a network or system to both detect and deter potential attackers (Ferguson-Walter, 2021). According to (Georginaet al., 2023), the techniques operate on the principle of misleading adversaries, gathering intelligence on their tactics, and buying time for cybersecurity teams to respond effectively. In addition, Ge et al. (2019) further submitted that by introducing other decoy elements such as false information into an organization's digital environment to mimic real assets, services, or data can also be used to lure the attacker away from the actual network, and when an attacker interacts with these deceptive components, it triggers an alert, notifying security personnel of potential unauthorized access or malicious activity (Hu et al., 2021). This early detection is critical as it enables organizations to respond promptly and mitigate potential threats before they escalate (Ulagwu-Echefu et al., 2022).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue V May 2025

Deceptive elements often referred to as decoys, can take various forms, including simulated servers, network segments, login credentials, or data files (Ferguson-Walter, 2021). These decoys are designed to closely resemble their real counterparts, making them entice to attackers, and by engaging with these decoys, attackers unknowingly reveal their presence and intentions, thus allowing security teams to intervene effectively (Oluoha et al., 2021). Hence, through the utilization of decoys and deceptive components, these methods not only enable the early detection of threats but also furnish organizations with actionable intelligence for the enhancement of safeguarding their digital assets (Uchechi et al., 2022).

Types of deception techniques

There exist a variety of deception methods, including honeypot and honey token, deception network and endpoint, deceptive data, and content.

Honeypots and honey tokens

Honeypots and honey tokens are crucial elements of cybersecurity deception strategies, providing organizations with effective tools to identify, analyze, and deter malicious actors in their network environments (Islam et al., 2020). Honeypots simulate real systems, luring potential attackers with their authenticity and triggering alerts when intruders engage with them. In parallel, honey tokens are fake data pieces placed within the network, resembling genuine information, and their interaction with attackers is closely monitored for insights into adversary tactics (Hu et al., 2021). These deception techniques according to Pawlick et al (2019) serves multiple purposes which are early warning systems, swiftly detecting unauthorized access or malicious activity; divert potential attackers away from genuine systems and safeguarding critical assets from compromise.

Deception Networks and Endpoints

Deception networks involve the creation of fabricated network segments with counterfeit assets, while deception endpoints are fake devices alongside real ones (Pawlick et al., 2019). These elements attract attackers, allowing security teams to gain insights into the attacker's movements and objectives within the network (Harbor et al., 2021).

Deception networks emulate counterfeit network environments to attract attackers, thus enabling the detection and analysis of attacks while deflecting them from legitimate networks. Similarly, deception endpoints are manufactured devices designed to lure attackers, thus facilitating the identification and analysis of attacks while diverting them away from genuine endpoints (Oluoha et al., 2021).

Deceptive data and Content

Deceptive data and content represent a strategic cybersecurity approach where organizations intentionally introduce counterfeit or misleading information into their digital systems. This practice involves the creation of fictitious data elements, such as fabricated customer records or financial information that closely mimic genuine data (Lu et al., 2022). The key objective behind this deception is to act as a tripwire; any unauthorized access or manipulation of this deceptive content serves as a conspicuous red flag, signalling potential security breaches and unauthorized activities within the network (Lopez et al., 2024). This proactive approach not only aids in the swift detection of malicious actors but also provides an opportunity for security teams to monitor and analyze the adversary's actions, gather valuable threat intelligence, and initiate timely incident response measures to protect sensitive data and critical assets. In essence Ma and Li et al. (2023) revealed that deceptive data and content serve as a valuable layer of defence, adding an extra dimension to an organization's security posture by actively luring and exposing potential threats while safeguarding the integrity of genuine data.

Advantages of Deception Technique application

Table 1 presents the various advantages of deception techniques which can be adopted for the management of zero-day exploits (Lee and Park, 2024; Javadpour et al., 2024).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue V May 2025

Table 1: Advantages of Deception Techniques

Advantage	Description
Early Threat Detection	Deception elements like honeypots and deceptive data detect zero-day exploits before they reach critical systems, enabling timely mitigation.
Behaviour-Based Detection	Instead of relying on known signatures, deception techniques monitor attacker behaviour, making them effective even for previously unknown exploits.
Low False Positive Rates	Interactions with decoys are rare for legitimate users, so any interaction is highly likely to be malicious which further helps in improving detection accuracy.
Diversion of Attacks	By luring attackers away from real systems, deception minimizes the potential damage of a zero-day exploit.

By considering the advantages identified in Table 1, the management of zero-day attacks can be made feasible and scalable due to the adaptive nature of deception techniques, hence, the application of deception techniques for the management of zero-day attacks, vulnerabilities and threats is recommended in future studies.

FINDINGS

The systematic literature review and analysis of the various studies on zero-day attack detection, deception security strategies and machine learning applications revealed several critical observations as presented in Table 2:

Table 2: Summary of Observations from the Review

S/N	Observation	Description	Implications
1	Rising adoption of deception technologies	Techniques like honeypots, honeynets and camouflaging are usually and increasingly used to detect, delay and mislead attackers.	Improves proactive threat detection and attacker profiling; should be incorporated into layered defence strategies.
2	High effectiveness of ML in anomaly detection	Models like autoencoders, CNNs, RNNs and SVMs show strong performance in detecting unknown threats with high accuracy.	Machine learning offers a scalable and adaptive approach for identifying zero-day attacks before signature-based approach can.
3	Under-explored ML-deception integration	Few studies combine deception strategies with machine learning for zero-day management.	There is significant research and practical potential in developing hybrid models that combine deception with intelligent analysis.
4	Dependence on static or outdated datasets	Datasets like CICIDS2017 and NSL-KDD are commonly used but may not fully represent modern threats.	The creation of dynamic, deception-generated or IoT-specific datasets is essential for model generalization.
5	Diversity of zero-day threats	Zero-day threats include malware, ransomware, adversarial ML attacks, and more.	Requires domain-specific threat models and tailored mitigation strategies.
6	Challenges in patch timeliness	Factors such as required user privileges, attack complexity, and vendor delays affect patch deployment.	Emphasizes the need for early threat detection and containment before patches are released.
7	Emergence of autonomous defense systems	Reinforcement learning agents and simulation tools like CyberBattleSim are used for adaptive and real-time defence of ZDA.	Encourages the shift towards self-defending, intelligent cybersecurity systems.
8	Ethical and legal concerns around deception	Studies noted potential risks, including legal liability and user privacy issues when deploying deception tools.	Organizations must consider regulatory compliance, ethical boundaries, and transparency when using deception.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue V May 2025

From the review, despite the growing body of research on zero-day attack detection and deception security strategies, significant gaps remain, a major shortcoming lies in the limited integration of machine learning algorithms with deception-based techniques such as honeypots, honeytokens, and mimicry. While both domains individually show promise, their combined potential remains largely untapped. Additionally, while machine learning is widely used for detecting zero-day vulnerabilities, the threat of adversarial attacks targeting these models through techniques like model poisoning or evasionhas not been sufficiently addressed. Hence, there is need for more studies to be conducted in this area of study.

Data Collection

To validate the practicability of this work, data of zero-day attack was collected and applied to develop behavioural analytical model used for zero-day detection. The data source is UGRansomeware dataset, collected from Kaggle repository. The dataset contain 14 attributes spread across three different classes which are malicious, benign and normal packet. Table 3 presents the table of data description, while data source is in the data availability section.

Table 3: Data description

Column Name	Description	Data Type
Time	Timestamp of the network flow	DateTime
Protocol	Communication protocol used (TCP, UDP, ICMP)	Categorical
Flag	Connection status or flow flags (SYN, ACK, FIN)	Categorical
Family	Type of malware family or legitimate software family	Categorical
Clusters	Grouping of flows/packets for threat detection	Categorical
SeedAddress	Source address (IP/domain) of the flow	Text
ExpAddress	Destination address (IP/domain) of the flow	Text
BTC	Bitcoin transaction amount	Numeric (Float)
USD	Equivalent USD amount of BTC transactions	Numeric (Float)
Netflow_Bytes	Total bytes transferred in the network flow	Numeric (Integer)
IPaddress	IP address of the network entity	Text
Threats	Threats associated with the flow (malware, phishing)	Categorical
Port	Port number used in the flow	Numeric (Integer)
Prediction	Model prediction for the flow (malicious, benign and normal)	Categorical

Machine Learning and Training

This work selected Random Forest (RF) as the machine learning of choice and then train to develop behavioural analytical model for behavioural analysis in cyber security. The RF is an ensemble learning algorithm made of several decision trees. During training, it creates many decision trees using random samples of the dataset and random subsets of features, allowing each tree to learn different patterns. When making predictions, it takes the majority vote from these individual trees to classify zero-day attack. The collected data was processed through imputation and normalization, and then it was splitted into training, test and validation set. The RF was then trained with the data. The training process involved creating multiple decision trees based on random subsets of the training data and random subsets of features at each split in the tree. Each decision tree is grown independently, and once all trees are trained, RF aggregates their outputs to make final predictions by majority voting. The final output is the model for the detection of zero-day attack.

Performance evaluation metrics

The metrics used for the evaluation of the model performance are precision, recall, f1-score and accuracy. Precision measured the positive prediction of zero-day attack. Recall measures the actual positive zero-day instances correctly classified, the f1-score measures the harmonic mean between precision and recall, while accuracy measure the overall success rate of the model in correctly classifying zero-day attack and normal packets. Equation 1-4 mathematically defined the metrics.

$$Precision = \frac{TP}{TP+FP} \quad 1.0$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad 2.0$$

$$\text{F1-Score} = 2 * \frac{Precision*Recall}{Precision+recall} \quad 3.0$$

$$\text{Accuracy} = \frac{TP+TN}{Total\ prediciton} \quad 4.0$$

RESULTS AND DISCUSSION

This section presents the result of the model training process. This was evaluated considering precision, recall, f1-score and accuracy. Figure 4 presents the confusion matrix of the data distribution between the predicted class and true class, while figure 5 presents the confusion matrix of the true positive and false negative predictions respectively. The confusion matrix discussed the results of the model training process, showing TP for the different classes. Figure 4 showed that when 8403 features of normal packets were feed to the model, 8346 was correctly classified as class 0 which is the normal packet without attack, 49 of the features were wrongly classified as benign features while 8 features were incorrectly classified as malicious packets. In the class 1 (benign attack), 48 of the features were wrongly classified as normal packet, 13280 of the features were correctly classified as benign while 44 of the features were wrongly classified as malicious attack. In the malicious class, 5 of the features were classified wrongly classified as class 0, 30 of the features were classified as class 1 while 7998 were correctly classified as class 2. In figure 5, the percentage TP was calculated and also the percentage false positive, it was observed that for normal packet, 99.36% TP was recorded, for class 1 which is benign reported 99.41% and malicious packet recorded 99.35%.

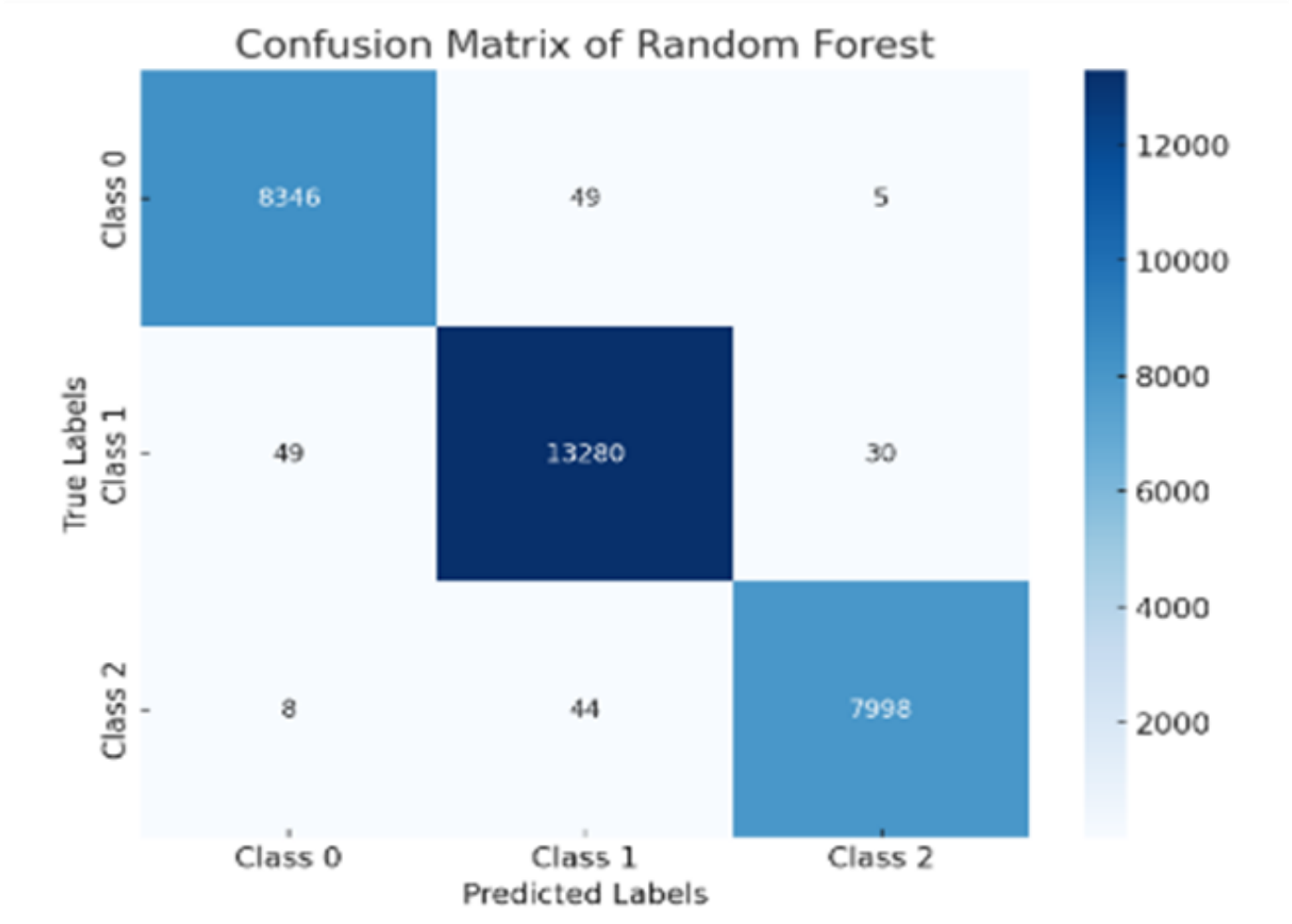

Figure 4: Confusion matrix of model

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue V May 2025

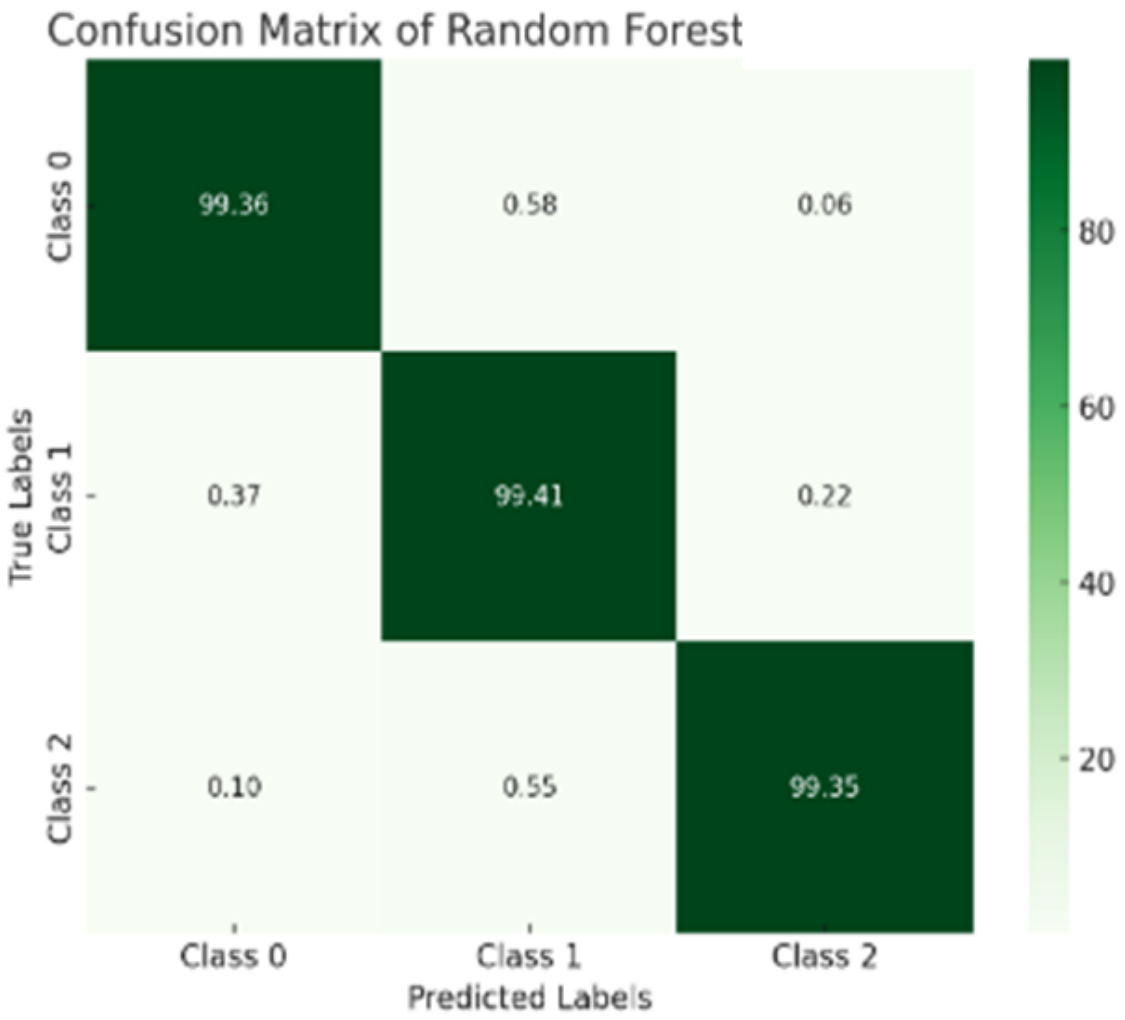

Figure 5: Confusion matrix of TP and FN

The figure 4 and figure 5 have presented and analyzed the performance of the trained RF model considering TP and FN. In the table 4, the results of the precision, recall, accuracy and F1 score was presented.

Table 4: Result of the RF training

Metrics	Precision	Recall	f1-score	Accuracy
0	0.99	0.99	0.99	
1	0.99	0.99	0.99	
2	1.00	0.99	0.99	
Average	0.99	0.99	0.99	0.99

From the result of the RF training, the precision recorded 0.99 which implied 99% of positive prediction of the classes of the dataset. Recall recorded 0.99 which implied that the model, was able to correctly classify actual instances of the data with 99% success rate. F1-score which measures the harmonic mean between recall and precision recorded 99%, while the accuracy of the model reported 99% success rate.

Comparative analysis with other state of the art zero-day management algorithms

In this section, a comparative analysis of existing model in literature tailored towards zero day attack management with our model was performed and all the results reported in table 5.

Table 5: Comparative analysis with existing algorithms

Author and year	Technique	Detection accuracy (%)
Manish et al. (2021)	Machine learning	98.00
Sharukh (2020)	CNN	87.50
Berna (2019)	DL	97.00
Ibraheem and Tasha (2024)	SVM	92.00
Ekong et al. (2023)	RF	95.00
Sarhan et al. (2023)	RF	87.37
	MLP	92.45
Our model	RF	99.0

Table 5 compare the performance of the model in correctly classifying zero-day attack considering different existing models and the new model developed with RF. From the results obtained, it was observed that our model recorded 99% accuracy and standout as the best for behavioural modelling of zero-day attack. The model was then recommended for integration with honeypot as decision based or classifying attacker to decoy network.

Data availability

The datasets used during the current study are available from the corresponding author on reasonable request. The UGRansome Dataset used in this study can be accessed via https://www.kaggle.com/datasets/nkongolo/ugransome-dataset.

CONCLUSION

This study conducted an extensive review on existing literature and emerging approaches in the domain of deception security for zero-day attack management. It explored the role of traditional deception techniques such as honeypots, mimicking and camouflaging, as well as their integration with Machine Learning (ML) algorithms to enhance threat detection capabilities. The study also examined how cyber-attacks exploiting network vulnerabilities such as zero-day vulnerabilities and threats pose a significant risk to system integrity, confidentiality and availability. The literature review presented in this study showed that ML techniques have proven effective in detecting anomalies and zero-day threats. The reviewed works equally demonstrated the potential of deception-based Intrusion Detection Systems (IDS) to mislead attackers and gather intelligence, yet many solutions proposed in this technique lack adaptability, real-time responsiveness and integration with advanced ML techniques. Additionally, the work presents the various advantages of deception techniques which can be adopted for the management of zero-day attacks such as early threat detection, behaviour-based analysis, low false positive rates and the ability to divert attacks away from critical systems. These features demonstrate the scalability, precision, and adaptability of deception mechanisms in complex cybersecurity environments. Experimental test was carried out with RF and results recorded 99% success rate in detecting zero-day attack. This model was recommended for integration with honeypot as deception solution for improved zero day management.

REFERENCES

1. Berna C., (2019) Zero-Day Attack Detection with Deep Learning. The Graduate School Of Natural And Applied Sciences Of Middle East Technical University.
2. Cen, M., Deng, X., Jiang, F., & Doss, R. (2024). Zero-Ran Sniff: A zero-day ransomware early detection method based on zero-shot learning. Computers & Security, 142, 103849. https://doi.org/10.1016/j.cose.2024.103849
3. Ekong A., Etuk A., Inyang S., & Ekere-Obong M., (2023) Securing Against Zero-Day Attacks: A Machine Learning Approach for Classification and Organizations' Perception of its Impact. Journal of Information Systems and Informatics Vol. 5, No. 3, September 2023 e-ISSN: 2656-4882 p-ISSN: 2656-5935https://doi.org/10.51519/journalisi.v5i3.546

4. Eze, E. M., Ituma, C., Asogwa, T. C., & Ebere, U. C. (2022). Development of machine learning based security algorithm for 4G network against wormhole. International Journal of Research and Innovation in Applied Science (IJRIAS), 7(2), 70–75.
5. Ferguson Walter, J. K., & Majo, K. M. (2021). Examining the efficacy of decoy-based and psychological cyber deception.
6. Ferguson-Walter, K., Fugate, S., Mauger, J., & Major, M. (2019). Game theory for adaptive defensive cyber deception. In Proceedings of the 6th Annual Symposium on Hot Topics in the Science of Security (pp. 1–8).
7. Ge, X., Han, Q.-L., Zhong, M., & Zhang, M.-X. (2019). Distributed Krein space-based attack detection over sensor networks under deception attacks. Automatica, 109, 108557. https://www.elsevier.com/locate/automatica
8. Georgina, C. S., Sakinah, F., Fadholi, M. R., Yazid, S., &Syafitri, W. (2023). Deception based techniques against ransomwares: A systematic review. Jurnal Teknik Informatika (JUTIF), 4(3), 529–553.
9. Guo, Y. (2023). A review of machine learning-based zero-day attack detection: Challenges and future directions. Computer Communications, 198, 175–185. https://doi.org/10.1016/j.comcom.2022.11.001
10. Harbor, M. C., Eneh, I. I., & Ebere, U. C. (2021). Nonlinear dynamic control of autonomous vehicle under slip using improved back-propagation algorithm. International Journal of Research and Innovation in Applied Science (IJRIAS), 6(9).
11. Hindy, H., Atkinson, R., Tachlatzis, C., Colin, J., Bayne, E., &Bellekens, X. (2020). Utilising deep learning techniques for effective zero-day attack detection. Electronics, 9(10), 1684. https://doi.org/10.3390/electronics9101684
12. Hu, Z., Deng, F., Su, Y., Zhang, J., & Hu, S. (2021). Security control of networked systems with deception attacks and packet dropouts: A discrete-time approach. Journal of the Franklin Institute, 358, 8193–8207. https://www.elsevier.com/locate/jfranklin
13. Ibraheem I., & Tasha A., (2024) Zero Day Attack Vulnerabilities: Mitigation using Machine Learning for Performance Evaluation. Journal of Computers for Society 5(1) (2024) 43-58https://doi.org/10.17509/jcs.v5i1.70795
14. Islam, M., & Al-Shaer, E. (2020). Active deception framework: An extensible development environment for adaptive cyber deception. In 2020 IEEE Secure Development (SecDev). https://doi.org/10.1109/SecDev45635.2020.00015
15. Javadpour, A., Jafari, F., Taleb, T., Shojafar, M., & Benzaid, C. (2024). A comprehensive survey on cyber deception techniques to improve honeypot performance. Computers & Security, 140, 103792. https://doi.org/10.1016/j.cose.2024.103792
16. Kovářová, M. (2024). Exploring zero-day attacks on machine learning and deep learning algorithms. In Proceedings of the 23rd European Conference on Cyber Warfare and Security (ECCWS 2024).
17. Kumar, R., & Subbiah, G. (2022). Zero-day malware detection and effective malware analysis using Shapley ensemble boosting and bagging approach. Sensors, 22(7), 2798. https://doi.org/10.3390/s22072798
18. Lee, S. Y., & Park, J. M. (2024). Sampled-data stabilization for networked control systems under deception attack and the transmission delay. Communications in Nonlinear Science and Numerical Simulation, 131, 107817. https://www.elsevier.com/locate/cnsns
19. Lopez, P. B., Nespoli, P., & Perez, M. G. (2024). Cyber deception reactive: TCP stealth redirection to on-demand honeypots. arXiv preprint. https://arxiv.org/abs/2402.09191
20. Lu, H., Wang, X., Zhou, W., & Gou, Y. (2022). Hybrid-driven-based H∞ filtering for networked systems under randomly occurring deception attacks. Journal of the Franklin Institute, 359, 6544–6566. https://www.elsevier.com/locate/jfranklin
21. Ma, Y., & Li, Z. (2023). Neural network-based secure event-triggered control of uncertain industrial cyber-physical systems against deception attacks. Information Sciences, 633, 504–516. https://www.elsevier.com/locate/ins
22. Manish A., Vivek S., Vishal K., Mahesh S., & Madhuri J., (2021) Detection of Zero-Day Security Threat Using Machine Learning. International Journal Of Current Engineering And Scientific Research (IJCESR).

23. Manish, A., Vivek, S., Vishal, K., Mahesh, S., & Madhuri, J. (2021). Detection of zero-day security threat using machine learning. International Journal of Current Engineering and Scientific Research (IJCESR).
24. Mohamed, N., Taherdoost, H., &Madanchian, M. (2024). Comprehensive review of advanced machine learning techniques for detecting and mitigating zero-day exploits. EAI Endorsed Transactions on Scalable Information Systems, 11(6).
25. Morozov, D., Vakaliuk, T., Yefimenko, A., Nikitchuk, T., &Kolomilets, R. (2023). Honeypot and cyber deception as a tool for detecting cyber attacks on critical infrastructure. DOORS-2023: 3rd Edge Computing Workshop, April 7, 2023, Zhytomyr, Ukraine.
26. Oluoha, O. U., Yange, T. S., Okereke, G. E., &Bakpo, F. S. (2021). Cutting edge trends in deception based intrusion detection systems—A survey. Journal of Information Security, 12, 250–269. https://doi.org/10.4236/jis.2021.124014
27. Oluoha, O. U., Yange, T. S., Okereke, G. E., &Bakpo, F. S. (2021). Cutting edge trends in deception based intrusion detection systems—A survey. Journal of Information Security, 12, 250–269. https://doi.org/10.4236/jis.2021.124014
28. Pawlick, J., Colbert, E., & Zhu, Q. (2019). A game-theoretic taxonomy and survey of defensive deception for cybersecurity and privacy. ACM Computing Surveys (CSUR), 52(4), 1–28. https://doi.org/10.1145/3338491
29. Peppes, N., Alexakis, T., Adamopoulou, E., &Demestichas, K. (2023). The effectiveness of zero-day attacks data samples generated via GANs on deep learning classifiers. Sensors, 23(2), 900. https://doi.org/10.3390/s23020900
30. Reddy, B., Shaik, S., & Gaddam, V. (2024). Reinforcement learning for zero-day vulnerability detection in IoT devices: A proactive approach. Research Square. https://doi.org/10.21203/rs.3.rs-4086508/v1
31. Roumani, Y. (2021). Patching zero-day vulnerabilities: An empirical analysis. Journal of Cybersecurity, 1–13. https://doi.org/10.1093/cybsec/tyab023
32. SakthiMurugan, S., Sanjay, K., Vishnu, V., & Santhi, P. (2023). Assessment of zero-day vulnerability using machine learning approach. EAI Endorsed Transactions on Internet of Things.
33. Sarhan M., Layeghy S., Gallagher M., & Portmann M., (2023) From zero-shot machine learning to zero-day attack detection. International Journal of Information Security (2023) 22:947–959 https://doi.org/10.1007/s10207-023-00676-0
34. Sarhan, M., Layeghy, S., Gallagher, M., & Portmann, M. (2023). From zero-shot machine learning to zero-day attack detection. International Journal of Information Security, 22, 947–959. https://doi.org/10.1007/s10207-023-00676-0
35. Sharukh S., (2020) A hybrid deep learning approach for detecting zero-day malware attacks.VNR Vignana Jyothi Institute of engineering and technology, Hyderabad, IndiaEasyChair Preprint
36. Sharukh, S. (2020). A hybrid deep learning approach for detecting zero-day malware attacks. VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India. EasyChair Preprint.
37. Singh, U., Joshi, C., & Kanellopoulos, D. (2019). A framework for zero-day vulnerabilities detection and prioritization. Journal of Information Security and Applications, 46, 164–172. https://doi.org/10.1016/j.jisa.2019.03.011
38. Teymourlouei, H., Stone, D., & Jackson, L. (2023). Identifying zero-day attacks with machine learning and data reduction methods. 2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE). https://doi.org/10.1109/CSCE60160.2023.00372
39. Tian, Y., & Zhao, N. (2024). Event-based adaptive secure asymptotic tracking control for nonlinear cyber–physical systems against unknown deception attacks. Journal of the Franklin Institute, 361, 106766. https://www.elsevier.com/locate/fi
40. Topcu, A. E., Alzoubi, Y. I., Elbasi, E., &Camalan, E. (2023). Social media zero-day attack detection using TensorFlow. Electronics, 12(17), 3554. https://doi.org/10.3390/electronics12173554
41. Uchechi, E., Ekeh, J., Eneh, I. I., Onoh, G. N., & Ebere, U. C. (2022). Enhanced transmission efficiency of multimode optic fiber for long distance data transmission using ANN controller. International Journal of Research and Innovation in Applied Science (IJRIAS), 6(12), 25–32.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue V May 2025

42. Ulagwu-Echefu, A., Eneh, I. I., & Ebere, U. C. (2022). Mitigating the effect of latency constraints on industrial process control monitoring over wireless using predictive approach. International Journal of Research and Innovation in Social Science (IJRISS), 5(11).
43. Walter, E., Ferguson-Walter, K., & Ridley, A. (2021). Incorporating deception into cyberbattlesim for autonomous defense. arXiv preprint. arXiv:2108.13980.