International Journal of Information Technology and Management
Vol. 19, Issue No. 1, February-2024, ISSN 2249-4510

DOI: https://doi.org/10.29070/md9c9378

Cyber Threat Intelligence in Intrusion Detection: A Systematic Review of Detection Strategies

Parag Deoskar [1] * , Dr. Ajay Kumar Sachan [2]

1. Research Scholar, LNCT University, Bhopal, M.P., India
deoskarparag@gmail.com ,
2. Professor, Dept. of CSE, LNCT University, Bhopal, M.P., India

Abstract: In the digital age, the need for robust Intrusion Detection Systems (IDS) is critical to safeguarding essential infrastructures due to the increasing sophistication of cyber security threats. While traditional IDS methods, such as signature-based and anomaly-based detection, have their merits, they often struggle to address emerging cyber threats like zero-day attacks, polymorphic malware, and advanced persistent threats (APTs). Recent advancements in machine learning (ML) and deep learning (DL) have significantly enhanced IDS capabilities, enabling them to detect threats in a more intelligent and adaptive manner. This review paper provides a comprehensive analysis of various intrusion detection approaches, including traditional, hybrid, and next-generation methods. It explores how deep neural networks (DNNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers can be used to identify complex attack patterns. Furthermore, we examine the role of feature selection techniques, data preprocessing methods, and publicly available datasets, such as UNSW-NB15, TCP/IP, and KDD99, in boosting the performance of IDS. The paper also discusses the challenges involved in implementing real-time IDS, including computational overhead, false positives, adversarial attacks, and scalability issues in cloud and IoT environments. Special attention is given to the potential of federated learning and blockchain-based IDS solutions for decentralized and privacy-preserving threat detection. Overall, this study provides researchers and cybersecurity professionals with a thorough understanding of the current state of intrusion detection, highlighting its limitations and potential advancements. The goal is to guide the development of more efficient and intelligent IDS solutions in the future.

Keywords: Intrusion Detection System (IDS), Network Security, Anomaly Detection, Machine Learning (ML), Deep Learning (DL), Cyber Threat Detection

---------------------------- X ----------------------------

INTRODUCTION

As networking technologies rapidly advance and cyber threats continue to rise, ensuring effective cybersecurity has become a top priority. Detecting and preventing unauthorized access and malicious activities within computer networks is a crucial aspect of cybersecurity. Intrusion Detection Systems (IDS) play a key role in monitoring network traffic and identifying potential security vulnerabilities. Traditional IDS methods often rely heavily on signature-based approaches, which, while useful, may fail to detect new and complex threats. To overcome these limitations, researchers and cybersecurity professionals have begun exploring the integration of machine learning (ML) techniques in IDS design. Machine learning has already proven successful in various domains, such as natural language processing, computer vision, and pattern recognition. This opens up promising opportunities to enhance the accuracy and effectiveness of intrusion detection systems in network cybersecurity. ML-based IDS can learn from large volumes of network data, detect anomalous patterns, and adapt to new attack strategies. This approach not only reduces false positives but also helps in identifying previously unknown threats. Building a machine learning-driven intrusion detection system (IDS) for network security involves several critical components. First, a strong and diverse dataset is needed to train and test the ML models. The dataset should include both normal and malicious network data to ensure effective learning. Second, appropriate feature selection and extraction techniques are essential to extract relevant information from the network data. These features enable ML algorithms to distinguish between regular traffic and potential intrusions. Moreover, the choice of ML algorithms plays a significant role in the design of IDS. Various algorithms, such as Support Vector Machines (SVM), Random Forests, and deep learning architectures, have been explored for intrusion detection. Each algorithm has its own advantages and limitations, and the selection of the most suitable one depends on factors like the complexity of the problem, the volume of available data, and the importance of detection accuracy and computational efficiency. This paper explores how to design and develop machine learning-based intrusion detection systems to protect networks. We review different machine learning techniques and methods used for both binary and multi-class classification tasks within IDS, assessing their strengths and weaknesses. Additionally, we discuss their performance across different network environments and attack scenarios.

Motivations

Intrusion Detection Systems (IDS) play a crucial role in safeguarding computer networks by identifying and preventing security threats. The use of machine learning models in IDS is growing due to their ability to process large datasets and detect patterns in real-time. These models enable IDS to learn from historical data and recognize emerging trends that might indicate potential intrusions, thus enhancing detection accuracy and reducing false alarms. This becomes particularly important when dealing with diverse datasets, as different datasets may exhibit unique characteristics, such as varying types of intrusions, network configurations, and user behaviors. By training machine learning models on such varied datasets, the IDS can adapt to different network environments and continue to identify threats across a broad range of contexts.

With the widespread use of networked computers and constant internet connectivity, which are vital to modern society, there are growing concerns about security. As internet-based technologies proliferate, attackers can now target computer systems remotely, posing risks to availability, privacy, and data integrity. Network traffic consists of packets with characteristics like duration, protocol type, and the volume of data exchanged between source and destination. Attackers may tamper with these packets during creation or transmission, making it essential to detect such attacks and ensure service integrity, as no networked system can be entirely immune to threats.

As network traffic volume increases, so do the number of unusual events, such as improperly configured devices, port scans used to prepare for future attacks, viruses and worms that consume resources and spread, and denial-of-service (DoS) attacks that disrupt network services. Accurately detecting and diagnosing these anomalies is critical to maintaining system functionality, especially in safety-critical systems (SCS), where failures can result in severe consequences, including injury, loss of property, or environmental damage. SCS encompass various applications, including industrial robotics, logistics, autonomous vehicles, medical systems, and security.

With the ongoing advancement of software-defined, self-driving, and interconnected systems, the risk of cyberattacks and their impacts continues to rise. As such, the ability to detect intrusions and abnormal behaviors is paramount for SCS. Some studies explore the use of IDS to monitor unusual communication flows in control systems within mechatronics and industrial applications. Unlike traditional attack monitoring algorithms in communication networks, which often involve complex, dynamic models of normal behavior, this method combines model-based and data-driven approaches. It utilizes state observers to identify discrepancies between the model and actual process data, providing a robust solution for anomaly detection.

This paper focuses on the security of communication networks by examining three commonly used datasets in this field. While significant progress has been made in designing machine learning-based IDS for network security, several research gaps and challenges remain. One major issue is the lack of labeled datasets suitable for training and testing IDS models. Many publicly available datasets are insufficient in size, lack diversity in attack scenarios, or contain outdated data, which hinders the development of generalized IDS models capable of detecting new and complex threats. Another challenge is the opacity of machine learning-based IDS models, particularly deep learning algorithms, which often function as "black boxes." This lack of interpretability makes it difficult to understand how decisions are made, reducing the trust and adoption of these systems, especially in critical security applications that require transparent explanations of detected threats. Moreover, much of the existing research focuses only on binary classifications normal versus malicious network traffic limiting the scope of IDS models.

THREATS, VULNERABILITIES, EXPLOITS, AND ATTACKS

In this part of the study, well-known cyber threats, security risks, weaknesses, and attacks are talked about in great detail. Cyber dangers are first looked at in terms of viruses, Trojans, worms, rootkits, and hackers. Next, well-known threats like ransomware, spyware, scareware, and joke programs are talked about. Next, security holes and the most common tools used to scan for them are shown. Last but not least, the most common types of attacks are talked about in detail. These include applications, cryptography, hacking, computer networks, phishing, malware, bots, botnets, password and man-in-the-middle attacks, and social engineering attacks. It also has suggestions, safety measures, and information about each type of attack. [13]

Network Security

This section provides a more detailed examination of the network security issue. It begins with a discussion of the Open Systems Interconnection (OSI) model, its various layers, and the protocols that operate at each layer. Following this, the security threats targeting each specific layer are explored in greater depth. The next section focuses on the various network security devices and tools available. Lastly, it covers the security of wireless networks, the types of attacks these networks face, and the strategies for safeguarding them.

The key contributions of this study, which are derived from a thorough exploration of the aforementioned topics, are summarized below:

1. The current state and problems of cyber security, as well as the latest technical advances in this area, are described;

2. Basic knowledge about the basics of cyber security is given;

3. Cybersecurity risks, dangers, attacks, and new research in these areas are shown;

4. Network security, OSI levels, and attacks on each layer are talked about.

5. Challenges, problems, and new ideas are put forward.

Considering the mentioned contributions, this study has some advantages for re- searchers in this field:

1. It is clear how important cyber security is;

2. Anyone, even a beginner researcher, can learn the basics of cyber security;

3. It is easy to find out about the most common threats and attack types;

4. It is possible to learn about the newest studies in this field; and

5. It is possible to get a comprehensive look at the cyber security field for further research.

Types of Cyber Security

One thing that each company has is assets, which are really just multi-system combinations. A strong cyber security posture is required for these systems, which necessitates coordinated efforts across all of its systems. Based on this, it is able to classify cyber security into the following sub-domains:

Network Security: This procedure includes installing the software and hardware required to secure a computer network against invasions, attacks, disruptions, misuse, and unauthorized access. With the help of this security, a company can more easily protect its assets from threats both inside and outside the company.

Application Security: It comprises taking precautions to prevent unwanted threats to the software and electronic equipment. Making sure the applications are always up-to-date and secure against attacks is one approach to accomplish this. Effective security is established throughout the design process, which includes activities like as coding, validation, and threat modeling, among others. This occurs immediately prior to the initiation of a device or application.

Information or Data Security: The implementation of a robust data retention system is required in order to preserve the confidentiality and integrity of data while it is being stored as well as while it is being transferred.

Identity management: The process of identifying the degree of access afforded to each individual possesses inside an organization is the subject of discussions in this section.

Operational Security: Performing this task requires processing and making judgments regarding the management and protection of digital assets.

Mobile Security: In the context of portable electronic devices like smart phones, laptops, tablets, and the like, it is necessary to protect the personal and organizational data that is saved on these devices from a variety of hostile attacks. These dangers include illegal access, the loss or theft of a device, malware, and other similar hazards.

Cloud Security: Safeguarding the company's data stored in the cloud or other digital environments is an integral aspect of this procedure. It protects itself from a lot of different kinds of risks by using a bunch of different cloud service providers, like Google, Microsoft Azure, and AWS.

Disaster Recovery and Business Continuity Planning: The focus here is on the procedures, checks, and warnings that a company puts in place to react appropriately in the case that any hostile activity results in the loss of data or operations. The company's policies require that it resume operations to the same capacity as before the crisis occurs, regardless of whether or not the activities were interrupted.

User Education: If any harmful action is leading to the shutdown of activities or the corruption of data, it is concerned with the protocols, monitoring, alarms, and plans that an organization employs in order to respond to the situation. In the event that a disaster occurs, its policies require that operations be resumed to the same level of capacity as they were before the occurrence.

IMPORTANCE OF CYBER SECURITY IN MODERN NETWORKS

At this point in time, we are living in a digital era, in which every facet of our existence is dependent on the network, different electronic gadgets, and software programs. Devices linked to the internet are fundamental to the functioning of every critical infrastructure, including healthcare systems, government agencies, banks, and manufacturing companies. They carry out their tasks with the help of this equipment. Private information, financial records, and ideas are all examples of the kinds of things that could be considered sensitive and so vulnerable to unauthorized access or disclosure, which could have disastrous consequences. [14] These details provide invaders and threat actors with the ability to enter them for the purpose of monetary gain, extortion, political or social motivations, or even simple vandalism. Hacking systems and other forms of cyber attacks might put the economy of the entire world in jeopardy. That is why cyber attacks have become a worldwide concern. In light of this, protecting vital data from highly publicized security breaches requires a robust cyber security strategy. Furthermore, with the ever-increasing frequency of cyber attacks, it is imperative that organizations and businesses, especially those dealing with sensitive data related to create rigorous cyber security policies and processes to safeguard sensitive information, such as health records, financial information, or national security documents.

Attack Detection Using Deep Learning Structures

Deep learning has been put to considerable use in the field of cyber security as a direct result of the amazing potential that it possesses for the development of various security applications. Associated applications include things like malware, infiltration, phishing, the detection of spam, and traffic analysis [15]. The examples of successful applications that have been offered here, in our opinion, could be of assistance in analysing the requirements of users with the innovation that has been brought about by deep learning architectures. As a consequence of this, in order to demonstrate that the deep learning method may be applied in a variety of contexts, we will provide a few instances of typical uses. Following is a description of certain programmes that, in our opinion, have the potential to be used in areas such as the management of multimedia files, signal processing, and other related areas. The applications of attack detection using deep learning structures are diverse and impactful, revolutionizing the landscape of cyber security.[8-13] Deep learning techniques have demonstrated remarkable capabilities in identifying complex attack patterns and anomalies across various domains. Some notable applications include:

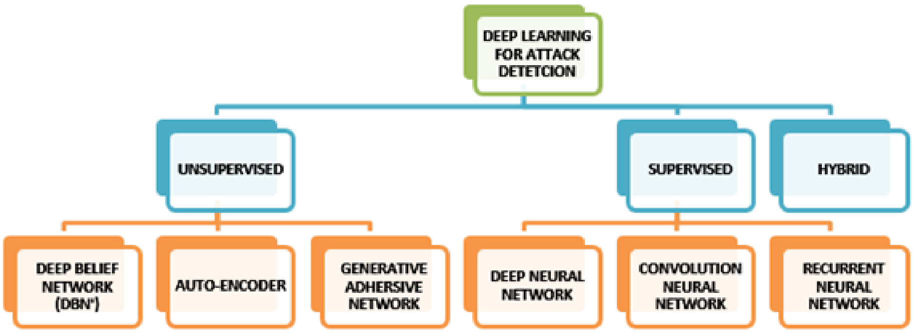
DEEP LEARNING FOR ATTACK DETETCION
UNSUPERVISED
SUPERVISED
HYBRID
DEEP BELIEF NETWORK (DBN')
AUTO-ENCODER
GENERATIVE ADHERSIVE NETWORK
DEEP NEURAL NETWORK
CONVOLUTION NEURAL NETWORK
RECURRENT NEURAL NETWORK

Figure 1: Categorization of the current deep learning methods for attack detection.

As was already said, critical assets are very important to the way society works. So, countries try to get better at what they do while also cutting down on time and money spent on production. Some of the changes that countries are making to reach this goal, though, have made key infrastructure systems less safe. Therefore, these holes in security let attackers get around those systems' defenses and get into them with rights they aren't supposed to have. briefly about some types of intrusion attacks below.

Brute-force attack: This strike is one of the best known. This type of attack is easy to carry out and depends on how powerful the attacker's computer system is. To be more specific, the attacker searches through all possible combinations of usernames and passwords to find the right one. Most of the time, these hacks use tools to figure out which letters and numbers make up the username and password. A Dictionary attack is one type of this type of attack. It looks up the right username and password by searching a dictionary for popular words and phrases that could be the username or password.

Buffer overload: The goal of this type of attack is to erase all the data in memory so that the attacker can take over the machine. For more specifics, the attacker feeds a computer more data than the buffer can handle. So, the data went past the buffer limit and the extra data that was stored in memory locations next to it. This is an attack that attackers use to either stop a Denial of Service (DoS) attack or, if the memory is clear, find the part of memory that stores the system's executable code and change it with their own. In the second case, the attacker can take over the server and change how the program works.

Phishing: Using phishing, scammers get people to give them their personal information by pretending to be someone else, like a company, that the person should trust [16]. People who want to attack usually send emails that look real but aren't. The user's email has a malware file or a link that isn't safe. So, the attacker wants the user (victim) to click on the link or document so that the malware can be put on the victim's computer. The attacker now has access to private information like passwords, usernames, and other things because of the last step. The following are some of the most popular types of phishing, as mentioned in [17]:

- Clone phishing

- Spear phishing
- Social networking on mobile
- Gaming phishing
- DNS base phishing
- Live chat
- Whalin
- Filter evasion

SQL injection: Set a SQL statement as an input in the application's input box to get into the database. This is an example of an intrusion strike. If the attack works, the hacker gets access to run malicious code that hurts the database or, even worse, gets control of private data. It goes into great depth about the risks of SQL injection attacks and how to classify the different types of SQL injection attacks.

Sniffer attack: Packages are often sent over the transmission route by applications to share information with each other. The attacker collects, decodes, inspects, and makes sense of the data in these packages. This is called a Sniffer attack. This kind of data generally includes passwords, usernames, and other important things. The Sniffer attack comes in two forms: active and silent. In an active Sniffer attack, the attacker changes network traffic. The target can tell if someone is spying on them and steal their packages. This is different from a passive Sniffer attack, where the attacker does not change network data and the victim does not know if someone is doing a Sniffer attack.

Trojan horses: As a way to get around the victim's system's defenses, Trojan horses hide the attacker's malware file inside other programs. Attackers often use these programs, which can be linked to emails or found in free downloads, to put malicious code on the victim's computer or do any other kind of attack they want. An attacker will often use trojan horses as back doors to get into a system.

LITERATURE REVIEW

Abbas Jamalipour (2021) Jamalipour introduces a network intrusion detection framework, HC-DTTWSVM, which employs hierarchical clustering for decision tree construction and twin support vector machines for efficient top-down intrusion detection. By leveraging benchmark datasets (NSL-KDD and UNSW-NB15), the study demonstrates the superior detection performance of the proposed model compared to newer methods, highlighting its capability to handle various types of network intrusion with minimized error accumulation.[18]

Xianwei Gao (2020) Gao addresses the challenges of network intrusion detection systems by proposing new evaluation metrics Detection Score and Identification Score for global performance comparison. Additionally, a workflow is presented to transform raw traffic into machine learning input features, facilitating systematic comparison across algorithms. The results affirm the simplicity and cost-effectiveness of raw traffic-based input for real-time, deep learning-powered IDS systems.[19]

MohammadNoor Injadat (2020) Injadat focuses on the IoT domain, highlighting the increasing risk of intrusions and proposing a hybrid IDS/IPS security framework. The study emphasizes artificial intelligence, particularly machine learning and deep learning techniques, to enhance detection and prevention. A comparative analysis identifies existing challenges and provides actionable insights for improving consistency and feasibility in security frameworks. [20]

Rui Fu (2023) Fu explores the application of deep learning in network intrusion detection systems, presenting a comprehensive study and categorization of frameworks. The survey highlights deep learning networks' effectiveness in intrusion detection and concludes with an analysis of IDS designs, providing observations and recommendations for future advancements in deep learning-based cyber security. [21]

Yeongje Uhm (2021) Uhm introduces XeNIDS, a novel framework for cross-evaluation of ML-based NIDS using network flows. The study utilizes labeled datasets for assessing wider use cases while addressing the complexity and risks of cross-evaluations. The proposed framework showcases how such evaluations can uncover unknown attributes of ML-based intrusion detection, improving their sensing scope without additional labeling costs. [22]

Chunyang Fan (2024) Fan investigates anomaly-based ML models for network intrusion detection, focusing on ensemble learning and convolution neural networks (CNNs). Using real-world data and the UNSW-NB15 benchmark dataset, the study demonstrates high detection accuracy, with CNNs showing slightly better performance, reinforcing their suitability for probing attack detection. [23]

Saikat Das (2021) Das proposes a multi-criteria, predictive cyber security framework for Industry 4.0, prioritizing and preventing breaches in wireless sensor networks. Using Decision Tree, MLP, and Autoencoder models, the framework achieves high accuracy in intrusion detection. Simulations indicate the MLP's superior classification performance, emphasizing its practical implications for safeguarding industrial IoT networks. [24]

Muaadh A. Alsoufi et al. (2021) presented a systematic literature review focusing on anomaly-based intrusion detection in IoT, highlighting that **supervised learning approaches are more effective** than unsupervised methods. The review analyzed seven deep learning approaches and concluded their utility in mitigating IoT security challenges, particularly in detecting zero-day attacks. [25]

Redhwan Al-amri et al. (2021) focused on the challenges and techniques for detecting anomalies in IoT data streams. The study covered important factors such as data complexity, evolving properties, and dimensionality. The authors emphasized the **need for innovative anomaly detection techniques** addressing these dynamic challenges, proposing a comprehensive picture of state-of-the-art methods. [26]

Muhammad Almas Khan et al. (2021) investigated a cross-layer IDS leveraging accumulated measure of fluctuation (AMoF) for classifying benign vs. malicious nodes in mobile networks. High performance (up to **98% detection rates)** was reported for high power/node velocity scenarios. [27]

Yakub Kayode Saheed et al. (2022) discussed **heterogeneity and energy efficiency** in IoT networks with self-organized nodes. Insights into lightweight frameworks for intrusion detection were explored, emphasizing cost and performance trade-offs. [28]

Vanlalruata Hnamte et al. (2023) describe the challenges introduced by the proliferation of IoT devices, focusing on the need for effective intrusion detection systems (IDS) in IoT networks. The study proposed a machine learning-based IDS using the UNSW-NB15 dataset. Their system demonstrated a high accuracy of 99.9%, showcasing its ability to identify attacks effectively using supervised machine learning models. [29]

Albara Awajan et al. (2023) and **Albara Awajan (2023)** highlight the rise of cyber-attacks against IoT due to its rapid expansion. They propose deep learning-based IDS, achieving promising results in detecting attacks such as DDoS, Sinkhole, and Blackhole with high detection rates—achieving 93.74% accuracy and a balanced precision-recall-F1 score in detecting intrusions. [30]

Iqbal H. Sarker et al. (2020) emphasize the vulnerability of IoT devices to attacks and present a deep learning-based intrusion detection method capable of detecting real-time malicious traffic targeting IoT devices, further bolstering efforts to secure these devices. [31]

Mohanad Sarhan et al. (2021) propose an intrusion detection model that uses machine learning, with an intrusion detection tree (IntruDTree) approach. This model reduces computational complexity and improves prediction accuracy while evaluating various feature extraction methods and their impact on detection accuracy. [32]

Abbas Jamalipour et al. (2021) discuss the challenges IoT networks face due to ineffective current IDS methods. The paper examines different machine learning models and feature reduction techniques, determining that a universal feature set is necessary for optimal detection performance across different datasets. [33]

Zahedi Azam et al. (2023) offer a comprehensive review on the IoT security landscape, categorizing various attacks and vulnerabilities. Their article also covers the application of intelligent intrusion detection strategies such as machine learning, deep learning, and reinforcement learning for improving IoT security. [34]

Table 1: different studies in network cyber security:

Author(s)	Year	Technique Used	Advantages	Disadvantages
Abbas Jamalipour	2021	HC-DTTWSVM (Hierarchical Clustering-based Decision Tree with Twin SVMs)	Superior detection performance, efficient intrusion classification	Potential scalability issues with larger datasets
Xianwei Gao	2020	New Evaluation Metrics (Detection Score & Identification Score)	Systematic comparison of IDS algorithms, cost-effective	Limited applicability beyond the tested frameworks
MohammadNoor Injadat	2020	Hybrid IDS/IPS using ML & DL	Enhanced security in IoT networks	High computational overhead

Rui Fu	2023	Deep Learning-based IDS	Improved intrusion detection accuracy	No real-time testing mentioned
Yeongje Uhm	2021	XeNIDS (Cross-Evaluation Framework for ML-based NIDS)	Identifies unknown attributes in ML-based IDS	Increased risk of overfitting
Chunyang Fan	2024	CNN-based Ensemble Learning for IDS	High detection accuracy, robust against probing attacks	CNNs require high computational resources
Saikat Das	2021	Multi-Criteria Cybersecurity Framework (Decision Tree, MLP, Autoencoder)	Effective for Industry 4.0 security	Not tested across multiple industrial settings
Muaadh A. Alsoufi et al.	2021	SLR on Anomaly-based IDS for IoT	Highlights supervised learning effectiveness	Unsupervised methods not well explored
Redhwan Al-amri et al.	2021	Anomaly Detection for IoT Data Streams	Addresses data complexity and evolving threats	Lacks real-time testing for IoT environments
Nahida Islam et al.	2021	DNN-based IDS for MQTT-based IoT	High accuracy across multiple datasets	Struggles with multi-label classification
Muhammad Almas Khan et al.	2021	Cross-layer IDS using AMoF	High detection rates (up to 98%)	Performance may drop in low-power networks
Yakub Kayode Saheed et al.	2022	Lightweight IDS for IoT	Balances cost and performance trade-offs	Not evaluated for large-scale IoT deployments
Vanlalruata Hnamte et al.	2023	ML-based IDS using UNSW-NB15 dataset	High accuracy (99.9%)	Requires feature engineering for optimal results
Albara Awajan et al.	2023	Deep Learning-based IDS for IoT	High detection rates (93.74%) for DDoS, Sinkhole, Blackhole	Needs further optimization for real-time performance
Iqbal H. Sarker et al.	2020	Deep Learning-based IDS for IoT	Real-time malicious traffic detection	Not robust against adversarial attacks

Mohanad Sarhan et al.	2021	ML-based Intrusion Detection Tree (IntruDTree)	Reduced computational complexity	Limited evaluation datasets
Zahedi Azam et al.	2023	Review of ML, DL, RL for IDS in IoT	Comprehensive analysis of IoT security trends	No experimental validation of proposed strategies
Javed Asharf et al.	2020	Review of IDS Methodologies	Highlights IDS challenges and dataset overview	No new detection framework proposed

INTRUSION DATASET

TCP/IP

The dataset that needs to be audited was given. It has a lot of different simulated attacks that happened in a military network. By simulating a normal US Air Force LAN, it set up a way to get raw TCP/IP dump data for a network. Multiple threats were sent to the LAN, making it feel like a real place. There is a set amount of time between each TCP packet in a connection, during which data moves from a source IP address to a target IP address according to a clear set of rules. Also, each link is marked as either normal or an attack, and each attack type is very clear. [37]

Each connection record consists of about 100 bytes.For each TCP/IP connection, 41 quantitative and qualitative features are obtained from normal and attack data (3 qualitative and 38 quantitative features) .The class variable has two categories:

Key Features of TCP/IP Data

TCP/IP data is usually collected from network traffic and consists of several essential features that describe communication between hosts. Below is a tabular representation of important

TCP/IP features:

Key Features of TCP/IP Data: TCP/IP data is usually collected from network traffic and consists of several essential features that describe communication between hosts. Below is a tabular representation of important TCP/IP features:

Table 2: TCP/IP

Feature Type	Feature Name	Description	Significance

Packet-Level Features	Packet Size	Size of the packet in bytes.	Helps analyze data volume and detect anomalies.
	Timestamp	Time when the packet was captured.	Used for time-series analysis and attack detection.
	Source IP	IP address of the sender.	Identifies the origin of the traffic.
	Destination IP	IP address of the receiver.	Determines the target of communication.
	Protocol	The protocol used (TCP, UDP, ICMP, etc.).	Helps in filtering traffic by protocol type.
Flow-Level Features	Source Port	Port number of the sender.	Used to differentiate multiple connections from a host.
	Destination Port	Port number of the receiver.	Identifies services being accessed (e.g., HTTP - 80, DNS - 53).
	Flow Duration	Time taken for the entire communication session.	Helps in analyzing connection persistence and DoS attacks.
	Number of Packets	Total packets exchanged in the session.	Indicates the intensity of communication.
	Number of Bytes	Total data transferred in the session.	Helps identify large data transfers or potential data exfiltration.
TCP-Specific Features	TCP Flags	Control flags (SYN, ACK, FIN, RST, etc.).	Helps detect abnormal TCP behaviors like SYN flood attacks.
	Window Size	The amount of data a sender can send before receiving an acknowledgment.	Used for congestion and performance analysis.
	Sequence Number	Unique number assigned to each TCP segment.	Helps in tracking lost or reordered packets.

	Acknowledgment Number	Number confirming received data segments.	Ensures reliable communication in TCP.
IP-Level Features	TTL (Time to Live)	Number of hops before the packet is discarded.	Useful for detecting traceroute and network congestion.
	Fragmentation	Indicates whether the packet is fragmented.	Helps in identifying fragmentation-based attacks.
	DSCP (Differentiated Services Code Point)	Defines packet priority and QoS level.	Used for traffic prioritization.

UNSW-NB15-

The UNSW-NB15 dataset is a widely used benchmark dataset for network intrusion detection. It contains both normal and malicious traffic, with a diverse range of modern attack types. The dataset includes 49 features categorized into different groups, each contributing to identifying network threats. [38]

UNSW-NB15- https://www.kaggle.com/datasets/dhoogla/unswnb15/dataWhether the model can accurately detect attacks or not is determined by evaluating it on datasets. In the end, the results of any NIDS are dependent on the data set's quality.

Table 3: Types of features in UNSW-NB15

Feature Type	Description	Significance
Flow Features (Basic Features)	General flow characteristics like duration, protocol, and packet counts.	Helps in understanding basic network behavior.
Content Features	Information extracted from the payload, like HTTP requests and response codes.	Useful for identifying specific attack patterns within packet content.
Time Features	Timestamps and inter-arrival times of packets.	Helps in detecting anomalies based on timing behavior.

Statistical Features	Metrics such as mean, variance, and standard deviation of packet sizes and flow duration.	Useful for identifying abnormal traffic patterns based on statistical deviations.
General-purpose Features	Basic properties such as source/destination IP, port, and protocol.	Helps in identifying traffic sources and destinations, which is crucial for attack detection.
Application Layer Features	Features related to higher-layer protocols (e.g., HTTP, DNS).	Helps in detecting application-specific attacks like SQL Injection or DNS spoofing.
Label Feature	Classifies each record as Normal or Attack (with attack types).	Serves as the ground truth for machine learning models.

KDD 1999

Featuring 25,192 TCP/IP connections (observations), the KDD 99 dataset is derived from a simulated LAN environment that imitates a baseline US Air Force configuration. A diversified dataset was produced by intentionally subjecting this network to a variety of attacks in an effort to make it reflect real-life events. Here, "connection" is the exchange of TCP packets between two IP addresses, one from the source and one from the destination, in accordance with a predetermined protocol. The data transmitted, the time it takes to start and end the connection, and other characteristics characterize these links. It classifies connections as "normal" or "anomalous" based on how they behave. Roughly 100 bytes of data is associated with each reported connection. With 38 numerical features and 3 qualitative attributes, every TCP/IP connection yields a total of 41 features. "Normal" and "Anomalous" are the two possible values for the class variable that determines whether a connection is considered an intrusion. A comprehensive overview of the dataset's features is provided in Tables 1 and 2, along with brief explanations of each. "Basic," "Content-based," "Time-based and "Connection-based" are the categories into which the qualities fall. The tables further detail the categorization of each feature as either "Continuous" (C) or "Discrete" (D). https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

KDD Cup'99

Collection of data in 1999, DARPA used the dataset of recorded network traffic from 1998 to produce the KDD'99 data set. Each network connection is undergoing preprocessing to create 41 features. Table 1 shows that the KDD'99 dataset's characteristics may be categorized into four sets: Basic characteristics (#1-#9), Content Features (#10-#22), Time based traffic features (#23-#31), and host based traffic features (#32-#41). The dataset KDD'99 [11] has 4,898,430 records, which is relatively large compared to comparable datasets. According to Table 2, the four most common types of assaults are Denial of Service (DoS), Unauthorized Access from a Remote Machine (R2L), Unauthorized Access to Root (U2R), and Probe. The KDD'99 dataset has been processed using a plethora of data mining methods designed to detect breaches in network traffic. It is common practice to use the KDD Cup'99 data collection while creating an

IDS. According to statistical analysis, the KDD data set has two big issues that greatly affect the system's efficiency. The abundance of duplicate records in the KDD dataset is a big issue. Both the train set and the test set include around 78% and 75% duplicate records, respectively. If there are a lot of duplicates, learning approaches might stop using a large number of records and start using incomplete ones. The algorithm will so cease to learn records that are infrequent. U2R, R2L, and similar networks may be vulnerable to these recordings.[39]

Table 4: Features Description KDD 99

Feature Name	**Type**	**Description**
Duration	C	Length of the connection
Protocol-type	D	Type of protocol
Service	D	Network service at the destination
Flag	D	Normal or error status of the connection
Src-bytes	C	Number of data bytes from source to destination
Dst-bytes	C	Number of data bytes from destination to source
Land	D	1 if connection is from/to the same host/port; 0 otherwise
Wrong fragment	C	Number of "wrong" fragments
Urgen	C	Number of urgent packets

CONCLUSION

Intrusion Detection Systems (IDS) have become crucial in countering cyber threats in modern network environments. This review has explored various IDS methodologies, including signature-based, anomaly-based, and hybrid detection approaches, discussing their respective advantages and limitations. Signature-based IDS are effective at detecting known threats but struggle to identify zero-day attacks and advanced cyber threats. In contrast, anomaly-based detection, powered by machine learning (ML) and deep learning (DL), has demonstrated promising potential in recognizing previously unknown attack patterns. However, several challenges persist, such as high false positive rates, adversarial attacks, scalability issues, and limitations related to real-time processing, all of which hinder IDS performance. As network environments become more complex—particularly with the rise of cloud computing, IoT, and edge computing—there is a growing need for IDS solutions that are scalable, adaptable, and capable of handling large data volumes with minimal latency. Future research should focus on the development of hybrid IDS models that combine

ML/DL techniques with traditional rule-based systems to enhance detection accuracy while minimizing false positives. Additionally, improving feature selection methods, optimizing deep learning architectures, and incorporating real-time adaptive learning techniques will be essential to ensure IDS robustness against emerging threats. Furthermore, lightweight IDS solutions tailored for resource-constrained environments, such as IoT devices and edge computing platforms, are crucial for bolstering cybersecurity in distributed systems.

References

1. Musa, U.S.; Chhabra, M.; Ali, A.; Kaur, M. Intrusion Detection System using Machine Learning Techniques: A Review. In Proceedings of the 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 10–12 September 2020; pp. 149–155.

2. Aljabri, M.; Altamimi, H.S.; Albelali, S.A.; Maimunah, A.H.; Alhuraib, H.T.; Alotaibi, N.K.; Alahmadi, A.A.; Alhaidari, F.; Mohammad, R.M.A.; Salah, K. Detecting malicious URLs using machine learning techniques: Review and research directions.IEEE Access 2022, 10, 121395–121417.

3. Okey, O.D.; Maidin, S.S.; Adasme, P.; Lopes Rosa, R.; Saadi, M.; Carrillo Melgarejo, D.; Zegarra Rodríguez, D. BoostedEnML: Efficient technique for detecting cyberattacks in IoT systems using boosted ensemble machine learning. Sensors 2022, 22, 7409.

4. Htun, H.H.; Biehl, M.; Petkov, N. Survey of feature selection and extraction techniques for stock market prediction. Financ. Innov. 2023, 9, 26.

5. Bhuyan, M.H.; Bhattacharyya, D.K.; Kalita, J.K. Network Traffic Anomaly Detection and Prevention: Concepts, Techniques, and Tools; Springer: Berlin/Heidelberg, Germany, 2017.

6. Zhendong Wang; Yong Zeng; Yaodi Liu; Dahai Li (2021)"Deep Belief Network Integrating Improved Kernel-Based Extreme Learning Machine for Network Intrusion Detection" DOI: 10.1109/ACCESS.2021.3051074, Page(s): 16062 – 16091, 12 January 2021

7. Ghada Abdelmoumin; Danda B. Rawat; Abdul Rahman (2022)"On the Performance of Machine Learning Models for Anomaly-Based Intelligent Intrusion Detection Systems for the Internet of Things" DOI: 10.1109/JIOT.2021.3103829, Page(s): 4280 – 4290, 10 August 2021

8. Safa Otoum; Burak Kantarci; Hussein T. Mouftah (2019) "On the Feasibility of Deep Learning in Sensor Network Intrusion Detection" DOI: 10.1109/LNET.2019.2901792, Page(s): 68 – 71, 26 February 2019

9. Gustavo De Carvalho Bertoli; Lourenço Alves Pereira Júnior; Osamu Saotome; Aldri L. Dos Santos; Filipe Alves Neto Verri; Cesar Augusto Cavalheiro Marcondes (2021) "An End-to-End Framework for Machine Learning-Based Network Intrusion Detection System" DOI: 10.1109/ACCESS.2021.3101188, Page(s): 106790 – 106805, 30 July 2021

10. Ngan Tran; Haihua Chen; Jay Bhuyan; Junhua Ding (2022) "Data Curation and Quality Evaluation for Machine Learning-Based Cyber Intrusion Detection" DOI: 10.1109/ACCESS.2022.3211313, Page(s):

121900 – 121923, 03 October 2022

11. Eysed Mohammad Hadi Mirsadeghi; Hayretdin Bahsi; Risto Vaarandi; Wissem Inoubli (2023) "Learning From Few Cyber-Attacks: Addressing the Class Imbalance Problem in Machine Learning-Based Intrusion Detection in Software-Defined Networking" DOI: 10.1109/ACCESS.2023.3341755, Page(s): 140428 – 140442,12 December 2023

12. Shahneela Pitafi; Toni Anwar; I. Dewa Made Widia,Faculty of Vocational (2023)"Revolutionizing Perimeter Intrusion Detection: A Machine Learning-Driven Approach With Curated Dataset Generation for Enhanced Security" DOI: 10.1109/ACCESS.2023.3318600, Page(s): 106954 – 106966, 25 September 2023

13. Bing Gao; Bing Bu; Wei Zhang; Xiang Li (2021) "An Intrusion Detection Method Based on Machine Learning and State Observer for Train-Ground Communication Systems" DOI: 10.1109/TITS.2021.3058553, Page(s): 6608 – 6620, 19 February 2021

14. Liu, J.; Dong, Y.; Zha, L.; Tian, E.; Xie, X. Event-based security tracking control for networked control systems against stochastic cyber-attacks. Inf. Sci. 2022, 612, 306–321.

15. Zha, L.; Liao, R.; Liu, J.; Xie, X.; Tian, E.; Cao, J. Dynamic event-triggered output feedback control for networked systems subject to multiple cyber attacks. IEEE Trans. Cybern. 2021, 52, 13800–13808.

16. Qu, F.; Tian, E.; Zhao, X. Chance-Constrained H-infinity State Estimation for Recursive Neural Networks Under Deception Attacks and Energy Constraints: The Finite-Horizon Case. IEEE Trans. Neural Netw. Learn. Syst. 2022

17. Chen, H.; Jiang, B.; Ding, S.X.; Huang, B. Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives. IEEE Trans. Intell. Transp. Syst. 2020, 23, 1700–1716.

18. Abbas Jamalipour; Sarumathi Murali (2021) "A Taxonomy of Machine-Learning-Based Intrusion Detection Systems for the Internet of Things: A Survey" DOI: 10.1109/JIOT.2021.3126811, Page(s): 9444 – 9466, 10 November 2021

19. Xianwei Gao; Chun Shan; Changzhen Hu; Zequn Niu; Zhen Liu (2019) "An Adaptive Ensemble Machine Learning Model for Intrusion Detection" DOI: 10.1109/ACCESS.2019.2923640, Page(s): 82512 – 82521, 19 June 2019

20. MohammadNoor Injadat; Abdallah Moubayed; Ali Bou Nassif; Abdallah Shami (2020) "Multi-Stage Optimized Machine Learning Framework for Network Intrusion Detection" DOI: 10.1109/TNSM.2020.3014929, Page(s): 1803 – 1816, 07 August 2020

21. Rui Fu; Xiaojun Ren; Ye Li; Yongtang Wu; Hao Sun; Mohammed Abdulhakim Al-Absi (2023) "Machine-Learning-Based UAV-Assisted Agricultural Information Security Architecture and Intrusion Detection" DOI: 10.1109/JIOT.2023.3236322, Page(s): 18589 – 18598, 30 January 2023

22. Yeongje Uhm; Wooguil Pak (2021) "Service-Aware Two-Level Partitioning for Machine Learning-Based Network Intrusion Detection With High Performance and High Scalability" DOI:

10.1109/ACCESS.2020.3048900, Page(s): 6608 – 6622, 04 January 2021

23. Chunyang Fan; Jie Cui; Hulin Jin; Hong Zhong; Irina Bolodurina; Debiao He (2024) "Auto-Updating Intrusion Detection System for Vehicular Network: A Deep Learning Approach Based on Cloud-Edge-Vehicle Collaboration" DOI: 10.1109/TVT.2024.3399219, Page(s): 15372 – 15384, 10 May 2024

24. Saikat Das; Sajal Saha; Annita Tahsin Priyoti; Etee Kawna Roy; Frederick T. Sheldon; Anwar Haque (2021)"Network Intrusion Detection and Comparative Analysis Using Ensemble Machine Learning and Feature Selection" DOI: 10.1109/TNSM.2021.3138457, Page(s): 4821 – 4833, 27 December 2021

25. Muaadh A. Alsoufi,Shukor Razak,Maheyzah Md Siraj,Ibtehal Nafea,Ibtehal Nafea,Fuad A. Ghaleb,Faisal Saeed,Maged Nasser (2021) "Anomaly-Based Intrusion Detection Systems in IoT Using Deep Learning: A Systematic Literature Review" 2021, 11(18), 8383; https://doi.org/10.3390/app11188383, 9 September 2021

26. Redhwan Al-amri,Raja Kumar Murugesan,Mustafa Man,Alaa Fareed Abdulateef,Mohammed A. Al-Sharafi,Mohammed A. Al-Sharafi,Ammar Ahmed Alkahtani (2021) "A Review of Machine Learning and Deep Learning Techniques for Anomaly Detection in IoT Data" 2021, 11(12), 5320; https://doi.org/10.3390/app11125320, 8 June 2021

27. Muhammad Almas Khan,Muazzam A. Khan,Sana Ullah Jan,Jawad Ahmad,Sajjad Shaukat Jamal,Awais Aziz Shah,William J. Buchanan (2021) "A Deep Learning-Based Intrusion Detection System for MQTT Enabled IoT" 2021, 21(21), 7016; https://doi.org/10.3390/s21217016, 22 October 2021

28. Yakub Kayode Saheed , Aremu Idris Abiodun , Sanjay Misra c, Monica Kristiansen Holone c, Ricardo Colomo-Palacios c(2022) "A machine learning-based intrusion detection for detecting internet of things network attacks" Volume 61, Issue 12, December 2022, Pages 9395-9409,

29. Vanlalruata Hnamte, Jamal Hussain (2023) "DCNNBiLSTM: An Efficient Hybrid Deep Learning-Based Intrusion Detection System " Volume 10, June 2023, 100053,

30. Albara Awajan (2023) "A Novel Deep Learning-Based Intrusion Detection System for IoT Networks" 2023, 12(2), 34; https://doi.org/10.3390/computers12020034, 5 February 2023

31. Iqbal H. Sarker,Yoosef B. Abushark.Fawaz Alsolami.Asif Irshad Khan (2020)"IntruDTree: A Machine Learning Based Cyber Security Intrusion Detection Model" 2020, 12(5), 754; https://doi.org/10.3390/sym12050754, 6 May 2020

32. Mohanad Sarhan , Siamak Layeghy , Nour Moustafa , Marcus Gallagher , Marius Portmann (2021)"Feature extraction for machine learning-based intrusion detection in IoT networks" Volume 10, Issue 1, February 2024, Pages 205-216, https://doi.org/10.1016/j.dcan.2022.08.012

33. Abbas Jamalipour; Sarumathi Murali (2022)"A Taxonomy of Machine-Learning-Based Intrusion Detection Systems for the Internet of Things: A Survey" DOI: 10.1109/JIOT.2021.3126811, Page(s): 9444 – 9466, 10 November 2021

34. Zahedi Azam; Md. Motaharul Islam; Mohammad Nurul Huda (2023) "Comparative Analysis of

Intrusion Detection Systems and Machine Learning-Based Model Analysis Through Decision Tree" DOI: 10.1109/ACCESS.2023.3296444, Page(s): 80348 – 80391, 18 July 2023

35. Elhanashi, A.; Lowe Sr, D.; Saponara, S.; Moshfeghi, Y. Deep learning techniques to identify and classify COVID-19 abnormalities on chest X-ray images. In Proceedings of the Real-Time Image Processing and Deep Learning 2022; SPIE: Bellingham, WA, USA, 2022;Volume 12102, pp. 15–24.

36. Zheng, Q.; Zhao, P.; Wang, H.; Elhanashi, A.; Saponara, S. Fine-grained modulation classification using multi-scale radio transformer with dual-channel representation. IEEE Commun. Lett. 2022, 26, 1298–1302.

37. Elhanashi, A.; Gasmi, K.; Begni, A.; Dini, P.; Zheng, Q.; Saponara, S. Machine Learning Techniques for Anomaly-Based DetectionSystem on CSE-CIC-IDS2018 Dataset. In Applications in Electronics Pervading Industry, Environment and Society: APPLEPIES 2022;Springer: Berlin/Heidelberg, Germany, 2023; pp. 131–140.

38. Pisner, D.A.; Schnyer, D.M. Support vector machine. In Machine Learning; Elsevier: Amsterdam, The Netherlands, 2020;pp. 101–121.

39. Widodo, A.; Yang, B.S. Support vector machine in machine condition monitoring and fault diagnosis. Mech. Syst. Signal Process.2007, 21, 2560