Deep Analysis of Risks and Recent Trends Towards Network Intrusion Detection System

D. Shankar[1*], G. Victo Sudha George[2], Janardhana Naidu J N S S[3], P Shyamala Madhuri[4]
Research Scholar, Department of Information Technology,
Dr. M.G.R. Educational and Research Institute Chennai, Tamil Nadu 600095, India[1*]
Professor, Department of Computer Science and Engineering,
Dr. M.G.R. Educational and Research Institute Chennai, Tamil Nadu 600095, India[2]
Assistant Professor, Department of Computer Science and Engineering,
Vishnu Institute of Technology Bhimavaram, Andhra Pradesh, 534202 India[3, 4]
Assistant Professor, Department of Computer Science and Engineering,
Vishnu Institute of Technology Bhimavaram, Andhra Pradesh, 534202, India[4]

***Abstract*—In the modern world, information security and communications concerns are growing due to increasing attacks and abnormalities. The presence of attacks and intrusion in the network may affect various fields such as social welfare, economic issues and data storage. Thus intrusion detection (ID) is a broad research area, and various methods have emerged over the years. Hence, detecting and classifying new attacks from several attacks are complicated tasks in the network. This review categorizes the security threats and challenges in the network by accessing present ID techniques. The major objective of this study is to review conventional tools and datasets for implementing network intrusion detection systems (NIDS) with open source malware scanning software. Furthermore, it examines and compares state-of-art NIDS approaches in regard to construction, deployment, detection, attack and validation parameters. This review deals with machine learning (ML) based and deep learning (DL) based NIDS techniques and then deliberates future research on unknown and known attacks.**

***Keywords*—Network; dataset; communication; intrusion detection system; attacks; deep learning; machine learning**

I. INTRODUCTION

The Internet is ubiquitous everywhere that plays a vital role in daily activities. It almost pervaded every field of business activities, from retail to space research for storing and sharing confidential data [1]. A vital concern in businesses is to secure the network from external disturbances. The advancement of new communication technologies and services increases the number of interconnected devices [2, 3]. The escalation of Internet based systems also increases vulnerabilities and virtual attacks on the network. Cyber-crimes on the Internet are rising exponentially with the rapid growth of technology. The cyber-attacks are severe attacks on the network. Thus, security is a critical subject for the network, and a failure in security may cause misleading information. Thus the network cannot meet its goals with the improper IDS [4]. Moreover, communication in the network is affected by individual hackers and malicious behavior, which seriously impacts the system [5]. Different types of attacks occur in the network, such as wormhole attacks, denial of services (DOS), black hole attacks, and flooding attacks. All these attacks prevent the user from accessing network resources and services by dropping or adding unnecessary packets [6]. Additionally, variations in malicious software will increase the threats to the network through security breaches.

An intrusion detection system (IDS) has drawn considerable interest among academics since it protects the network against inside and outside attacks, a proactive ID tool. The IDS is a security application used after the antivirus software and firewall. The IDS alters the network when any faults occur in the system. The IDS not only detects intrusions but also stores the network traffic in the system [7]. The IDS alerts the system when the intrusion occurs at the host and network levels [8]. The available IDS have many limitations having difficulties coping with the cyber-attacks due to their dynamic nature. The IDS cannot be easily adaptable; thus, it cannot detect new malicious attacks [9].

Furthermore, the massive network size and the larger number of applications handled by the node have resulted in the IDS becoming a challenging task owing to the generation of huge amounts of data. The massive data generated in the network is communicated with other nodes inside the network will create new attacks [10]. Hence, several methodologies have been developed in the past few issues to deal with these issues.

Deep learning (DL) and machine learning (ML) approaches have recently been widespread in several applications. The DL based methods have functioned with the neural networks containing numerous layers. In ML, the algorithms analyze the data to solve the issues by learning and pattern discovery. These methods are found to be excellent in higher detection accuracy and better performance on dataset testing; hence, these methods are included in IDS. Thus this paper has reviewed traditional methods to ML and DL based methodologies for ID by considering different types of faults [11]. Recently, the DL and ML based approaches gaining more attention in the IDS context. In fact, the DL is used for feature extraction, selection and classification. The DL networks carry the learning process in supervised and unsupervised manners. The Internet is omnipresent in all fields, which is inevitable in network communication. Thus the applicability of IDS in networks is considered a vital

research topic for enhancing network security against attacks. However, several authors have made a review this topic. Some of the recent papers of review of IDS are listed below:

- A review on DL based IDS is conducted in [12, 13] that only examines the DL based methods in IDS, but that paper did not focus much on network attacks.
- Recently, the IDS in the Internet of Things (IoT) have been reviewed, and some solutions are suggested to overcome the issues [14].

The existing review papers provide a deep insight on DL, ML or existing approaches individually. However, they cloud not analyze up-to-date methods. Moreover, different types of datasets are not focused in deep. Apart from these recent reviews, this paper is intended to focus on the up-to-date methods for IDS thus it reviews traditional methods. The major aim of this paper is to review various IDS methods, ID behavior, performance efficacy and problems associated with IDS methods. Furthermore, the leading methods in the IDS are also discussed with their advantages. Although several techniques have been developed in this area, no single method covers all types of network attacks. In the paper, innovative content is added in terms of up-to-date methods by including all traditionally available methods for IDS in the network by considering different types of attacks in the network. Thus, this review has provided the growth of IDS and the issues the developed methods face.

The contribution of this paper is summarized as follows:

- To study the network security issues and point out the necessity of an IDS as a solution for detecting attacks.
- To emphasize attacks in networks, then provide the key role of IDS for network security.
- Highlight the challenges and advantages of the current works in intrusion detection.
- To bring an effective solution to overcome the issues associated with recently developed methods.
- Provide an excellent approach for detecting the harmful unknown intrusions in the network by resolving the issues in existing and newly developed methods based on the analysis.

The organization of this review paper is given as follows; section-2 deals with several attacks in the network, and Section 3 elaborates on the detection methods and their pros and cons. Section 4 explains the limitation of NIDS. Section 5 includes the conclusion and future scope for network ID.

II. Analysis of Attacks in the Network

Data transmission through the network may be subject to various threats and safety risks. Hence, it is often unable to be secure with new techniques. It is very important to analyze the intrusion and attacks in the network to secure it from failure. Malicious nodes in the network create a routing attack. Different security threats and network attacks are discussed below.

A. Security Threats in Network

The network should meet some security measures to prevent intrusions from outside the network. An unauthorized user can enter the network if the organization fails to meet sufficient security concerns like prediction, securing and isolation. On the other hand, data management in the network is affected by a lack of bandwidth management. Because data communication in the network depends on bandwidth resource utilization, thus the excess utilization of bandwidth in the network will lead to bandwidth wastage. The severe issue in network transmission is data confidentiality. The lack of an encryption mechanism will lead to data leakage in the network [15]. Network attacks such as snooping, traffic, modification, denial of services and injection of intrusions are common in the network [16].

B. Active and Passive Attacks

Attacks in the network are classified into two types: active and passive attacks. The intruder incepts the data in the network during a passive attack, whereas, in an active attack, the intruder initiates a command to disturb network operation.

- Spoofing: In a spoofing attack, the sender can change the network topology due to the miss-present data of a malicious node.
- Modification: In this kind of attack, the data route is modified by a malicious node so that the message may be sent through a long route. Data routing through long-distance leads to communication delays [17].
- Wormhole attack: In a wormhole attack, the data transmission between two nodes is affected due to malicious nodes. This attack fakes a route instead of the short route in the network; thus, distance gets confused [18].
- Sinkhole: The sink-hole attack attracts network traffic through false routing metrics. The sink-hole attack invites many attacks in the network, sending false information [19].

- Denial of service attack: Most communication networks have limited energy resources; hence, they do not cope with sophisticated safety technologies. One of the common attacks in network communication is a DOS attack. The DOS affects almost all network layers, thereby disabling the proper functioning of the network. The DOS also abuse data in the network by allowing multiple attackers. The DOS attacks may lead to poor network performance, spam messages, packet delay and loss [20].
- Sybil attack: This attack uses fake identities to transmit data on the network.
- Traffic analysis attack: this kind of attack examines node behavior, network traffic and length of the message.
- Eavesdropping: Eavesdropping on ongoing communication may cause information leakage on cryptography or connection [21].
- Monitoring attack: The intruder can read the confidential data but cannot modify the data.
- Remote to Local (R2L) attack: The attacker gets access to the network without an account in the R2L attack, considered a critical attack in the network. This type of attack depends on host level and network-level features.
- User to Root (U2R) attack: The U2R needed semantic data that was hard to capture at the initial stages. U2R commonly happens in content-based fields where the attacker starts as a normal user and then becomes a superuser to abuse the network.
- Probe: It is illegal to gather network information about several services and sources to violate network security.
- Man-in-the-middle (MITM) attack: In the MITM attack, the attacker can access both ends of the communication channel, thereby manipulating the data. In such attacks, the attacker tries to initialize secure communication by sending messages. Finally, the attacker gets access to encrypt all the messages in the communication channel [22].
- Malware attack: In this kind of attack, the evade signature is matched by dynamically modifying the code. Due to the different purposes of the attackers, it becomes difficult to detect the attack.
- Phishing attacks: Spam e-mails are created to advertise a product, which is slightly modified to harm the user is known as phishing attacks financially. Phishing crimes are created by making a fake website the same as the original website and then creating fraudulent offers to the user [23].
- Distributed denial of service: This attack disrupts the normal traffic of the targeted server or network by overwhelming the target.

III. Background of Intrusion Detection System (IDS)

Intrusions in the system are noticed by continuous observation of the system, in which the system administrator examines intrusions through user activities. The main aim of IDS is to examine the intrusions in the system. However, it is a difficult task, in fact, the IDS has not examined intrusions at all; that only examines the symptoms of intrusions. The symptoms gathered by the IDS are referred to as manifestation, where the collected evidence is not sufficient for detection, then the system can't access the intrusions at all. As the network attacks increased with time, the early form of intrusion detection methodologies is not scaleable.

The IDS is a significant tool for the network to examine security breaches. The IDS continuously monitors the network traffic entering and leaving the system, thereby examining network intrusions. The communication in the network is carried out through wire and wireless mediums. Thus, the network attacks increased dramatically in the network. Moreover, technological advancements increase new attacks on the network. The intrusion of attacks common in new computing environments such as wireless sensor network, fog computing, e-healthcare and cloud computing. IDS is vital security component that enables the computer network for IT organization. Hence, it is needy to construct IDS to defend the network from attacks. To provide a sufficient solution for IDS, it is important to investigate the up-to-date methods.

A. Principles of IDSs

IDS is nothing but the process of observing the events that occur in the system and then examining the intrusions through these events [24]. Intrusions can take several forms routing attacks, sniffer attacks, U2R attacks, man-in-the-middle attacks, Dos/DDos, and cyber-attacks [25, 26]. One of the major tasks of IDSs is to separate harmful attacks from normal traffic in an efficient manner. Connecting to a wrong system by wrongly typing the address is an example of normal traffic, which is considered a threat. Due to the timely detection of intrusions, IDS protects the system or network from failure. The major functions offered by IDS are listed below.

- Provide timely alarm while detecting threats in the system.
- Take necessary action to respond to the alarm detected [27, 28].

Factors to be considered for effective IDSs are summarized below:

- Robustness of the system
- Speed of detection
- Maximum detection rate.
- The minimum detection rate of normal traffic is a threat.
- Reduction in the system requirement, which includes hardware and software
- Accuracy in detecting threat location

- Integration with other technologies.

1) Data collection & recording: Initially, the network data are gathered to create the profiles of normal data compared with the observed data of the network. After initializing the profile of normal data, the network data are collected to check the intrusions. Since the volume of collected data is huge, creating small groups as vectors is useful in detecting intrusions.

2) Identification of harmful intrusions: Harmful threats in the network are observed by analyzing the data. In this step, the host and network data were examined for accurate detection of intrusions. In this stage, the undesired actions outside the network are recorded to detect intrusions.

3) Alert: In this stage, the condition of the network is examined by the judgement based on evaluated data, which examines whether intrusion occurred or not. Once it is acknowledged that the threat has happened, the IDS immediately communicate it with the administrator. Few IDSs can control attacks by utilizing network resources. Communication of threat occurrence is performed using various platforms like e-mails and messages in the user interface [29, 30].

IV. Detection Methodologies

The ID methodologies are classified a:

- IDS types by detection technique
- IDS types by monitored platform

A. IDS Types by Detection Technique

Further, the detection technique is categorized as,

- Anomaly based model
- Specification based
- Hybrid methods

1) Anomaly-based (anomaly detection): Anomaly-based IDS detection approaches compare observed activities with definitions to predict malfunctions. Normally, anomaly-based detection has some rules to define basic network functioning based on the intrusions observed. Sultan et al. [31] proposed an anomaly based method using a variational encoder for the ID in the network. The intrusion was detected through a semi-supervised learning approach and unsupervised deep learning (DL) methods. Variational autoencoder (VAE) and autoencoder (AE) were employed using flow features to detect unknown attacks. The area under receiver operating characteristics (ROC) was calculated using these methods and compared with a one-class (OC) support vector machine (SVM). AE and VAE were used, and OCSVM were trained through semi-supervised learning. To detect intrusion on flow based data, that approach uses DL methods. But that proposed method increases the false alarm rate.

Whereas, Shubhra et al. [32] proposed an adaptive scheme that combines both the adaptive grasshopper optimization algorithm (AGOA) and the ensemble of feature selection (EFS) method for identifying attacks. The EFS ranked the attribute from a selected subset of attributes having higher ranks, and the AGOA was adopted to find significant attributes from datasets. SVM was used as a fitness function to increase classification performance and efficient feature selection; the proposed method decreased proficiency. Hence to improve the classification, Bayu et al. [33] proposed a Two-Stage Classifier based IDS (TSE-IDS) to detect attacks. In order to minimize the feature size of datasets, hybrid particle swarm optimization (PSO), ant colony algorithm (ACO), and genetic algorithm (GA) were used. In addition, a reduced error pruning tree (REPT) classifier is used for improving classification performance and feature selection.

In order to improve dynamic network performance, Nguyen et al. [34] proposed anomaly-based network IDs (NIDS) using DL to consider false-positive rates and the unavailability of labelled data. The proposed method used Restricted Boltzmann Machines (RBM) and Autoencoder. The staked auto encoder performed better than RBM but consumed more time due to higher computation. Future work is needed to improve speed and reduce the oscillations in the slope of training error. Roshan Kumar and Deepak Sharma [35] proposed a hybrid ID algorithm focusing on reduced time consumption. That proposed method was developed by combining anomaly and signature-based approaches. That proposed method was more effective for detecting more attacks. But that proposed method had the drawback of higher time consumption.

TABLE I. Anomaly-Based IDS

Author	Method	Datasets	Outcome	Advantages	Drawbacks
Sultan et al. [31]	Variational Autoencoder	NSL-KDD	AUC – 0.7596	The detection rate is higher	A false alarm is more
Shubhra et al. [32]	Adaptive scheme	ISCX 2012	A-99.13 DR-99.23 FPR-0.067	predict the networks traffic behavior accurately	Decrease in proficiency
BAYU et al. [33]	Two-Stage Classifier Ensemble	NSL-KDD and UNSW-NB15	A - 85.8% S – 86.8% P-91.60% DR-88%	Improved precision metric and accuracy	Multi-class classification problem
Nguyen et al. [34]	NIDS	KDDCup99	T-240sec	Effective in detecting and classifying intrusion into five groups.	Higher oscillation in the slope of training error
Roshan et al. [35]	HyINT	KDDcup99	T-14.773	Ability to detect unknown attacks	Time-consuming process

Above mentioned methods for anomaly-based IDS methods for network ID and their outcome and advantages are illustrated in Table I. In every table, the performance measures are indicated by Accuracy (A), Detection rate (DR), Precision (P), F-score (F), Connection rate (CR), Learning rate (LR), True Positive rate (TP), False Positive rate (FPR) and Ability to avoid misclassification (AUC), false alarm rate (FAR), false positive (FP).

2) Specification-based NIDS: Specification-based IDS uses manual specifications to detect network intrusion. Various specification-based methods are listed below.

Anhtuan Le et al. [36] proposed Routing Protocol (RP) for Low Power and Lossy Networks (RPL). That proposed method was constructed by a semi-auto profiling technique that generates high-level abstracts. That proposed approach includes protocol states and transitions on statics executed on several IDS rules. The power consumption was minimized by eliminating the overhearing of communication. Furthermore, the RPL information object and Information Solicitation (DIS) were introduced to alleviate the synchronization issues. But that the proposed method needs to be improved to detect internal threats.

Hence, Herson Esquivel-Vargas et al. [37] proposed a specification-based IDS using the BACnet protocol to enhance the detection rate. In that approach, fully automated deployment of IDS through BACnet protocol was used. In that protocol, the certified devices were demanded to document, representing network behavior. The prototype of that protocol was executed passively, and the attacks were detected on a single BACnet packet. But that proposed method needs to be improved to increase network security.

Above mentioned methods for specification-based IDS, along with its outcome and advantages, are illustrated in Table II.

TABLE II. SPECIFICATION-BASED METHODS FOR IDS

Author	Approaches/ modules	Application	Outcome	Advantages	Drawbacks
Anhtuan Le et al. [36]	semi-auto profiling technique	RPL-Based Network Topology	OH-6.3%	Higher energy efficiency	An extension of IDS is needed for detecting internal threats
Herson Esquivel-Vargas et al. [37]	synthetic traffic	BACnet Protocol	P-99.85% R-99.57%	Attacks can be detected during real and synthetic traffic	System security needs to be improved

3) Hybrid techniques: Hybrid techniques are developed by combining two methodologies for ID. Thus, the hybrid methods overcome the drawbacks of single approaches. Various hybrid approaches for NIDS are given below.

Further improving network performance and addressing the issues of RPL, Areej Althubaity et al. [38] proposed an architecture by combining a centralized model with a sink and a distributed module with RPL. That hybrid method was called Authenticated Rank and Routing Metric (ARM). That proposed approach validates the legitimacy of data transferred through RPL control data while constructing the route. Sajad Einy et al. [39] proposed hybrid anomaly and signature-based NIDS for enhancing network security. Suricata IDS was adopted with a neural network model to examine the network's malicious behaviour in that proposed model. The signatures of different types of attacks were examined by the neural network (NN) model. Then, the output from NN was given to Suricata IDS. In addition, an open-source blacklist internet protocol was adopted in that system.

That proposed method may create additional attacks in the network. Hence to avoid additional errors in the network, MavraMehmood et al. [40] proposed a hybrid approach by combining SVM and an adaptive neuro-fuzzy interference system (ANFIS). That paper adopted min-max and data transformation methods for data pre-processing. Then, optimum features were chosen by the random forest recursive feature elimination technique. Afterwards, ANFIS and SVM were used for ID and classification, respectively.

Above mentioned hybrid methods for IDS, along with its outcome and advantages, are illustrated in Table III.

TABLE III. HYBRID METHODS FOR IDS

Author	Method	Approaches/ modules	Outcome	Advantages	Drawbacks
Areej Althubaity et al. [38]	hybrid specification-based IDS	Centralized and distributed models	Power consumption is 1966.711 mW	Extra overhead is minimized	-
Sajad Einy et al. [39]	Hybrid anomaly and signature-based IDS	Suricata IDS	A-96.11%	Different attacks are detected	Lower accuracy
MavraMehmood et al. [40]	ANFIS and SVM	random forest recursive feature elimination technique	A-99.3% Sp-0.998% P-0.999% R-0.992% F-0.995%	Higher accuracy	Higher cost

B. IDS Types by Monitored Platform (Data Source)

The monitor-based IDS methods are classified as:

- Network based IDS
- Host-based IDS

1) Network based IDS (NIDS): A NIDS detects harmful traffic on a network requiring promiscuous network access to examine the traffic. The IDS used some components to protect the network from attacks. The NIDS management console and management server were secreted from the remaining network. Therefore, the attacker cannot determine the location of the components.

Basant Subba et al. [41] proposed an artificial neural network (ANN) for IDS. The ANN optimization techniques

minimize the computational overhead and maintain high-level performance simultaneously. ANN consists of more interrelated nodes collaborating near each other to make a solution an exact task. ANN depend IDS model was suitable for real-time consumption. That proposed method requires more processing time. Hence to reduce the time consumption, Norbert Adam et al. [42] proposed NN based IDS to identify cruel behavior in the network. Ethernet taps were used to divide the signals and then distribute one branch to the original objective and another branch to the IDS. That technique tested for the nMap scanning attack, UDP flood attack, SYN flood attack and non-hateful statement.

Vrushali D.Mane & SN Pawar [43] proposed a back propagation ANN (BPANN) algorithm for detecting various network attacks to reduce FAR. The main objective of those methods was safeguarding the complete data with the support of a supervised neural network. In that approach, the neural network only used significant features of the KDD 99 dataset. The system performance was analyzed using 10% of the data from the KDD 99 dataset. The KDD training dataset contains a large number of signal connection vectors. The neural network needs a lot of time to test and train all datasets.

Further, Hossein Gharaee & Hamid Hosseinvand [44] proposed a GA and SVM for feature selection to improve network performance. GA drops the data dimension similarly, improving true positive detection and reducing the FP detection. In the feature selection method, input was considered traffic data that produces features (chromosomes) and then chooses chromosomes with higher categorization precision. SVM reduced computational time for training and achieved a low FPR with high accuracy. In that technique, KDD CUP 99 and UNSW-NB15 datasets were used for the testing. That proposed model was not suitable for large datasets.

IDS was a valuable device for analyzing and detecting malicious activities in the cloud network, and that was employed in organizations, enterprises and is important in cyberspace security. Kai Zhang et al. [45] proposed an intrusion action based correlation framework (IACF) for analyzing and correlating malicious behaviors in networks. IACF was used for improving the procedure of action extraction, scenario discovery and alerting. IACF alerts the network depending on the conception of intrinsic tough correlations. The sequence pruning algorithm (SPA) decreases false-positive impact and constructs the correlation. The IACF technique was used for predicting intrusion behavior depending on correlation graphs.

Basant Subba et al. [46] proposed a game theory (GT)-based false alarm (GTBFA) for reducing the FPA in signature-based IDS. A high FPA rate led to considerable utilization of network assets for monitoring against useless network fear. That proposed method links IDS alarms with network behavior to reduce the FPA rate of IDS. GTBFA uses more malicious activity scanners to scan the cloud network and generate a threat summary of the network. DARPA and IITG lab network datasets were used to reduce the FPA rate of IDS. A GT procedure was used to develop the network's sensible vulnerability set (SVS). This proposed method offer reduces the detection rate of critical vulnerabilities. At the same time, network classification had long-term problems such as irrelevant features and redundancy. Priyadarsi Nanda et al. [47] proposed the least square (LS) SVM for IDS for feature selection. The KDD Cup 99, NSL-KDD and Kyoto 2006+ datasets were used to examine the performance of the LSSVM based IDS. Flexible, mutual information feature selection (FMIFS) was an efficient feature selection algorithm to decrease similarity. SVM was able to solve the binary classification problems.

Mobile malware could direct to some cyber security threats such as installing backdoors, stealing sensitive information, sending premium SMSs and ransomware attacks. The antivirus systems were not able to detect the advanced threats. Therefore, there is an additional layer of safety on the network to protect the users from threats. Sanjay Kumar et al. [48] proposed a machine ML dependent on NIDS. ML classifiers were built using a dataset consisting of labeled instances of network traffic features created through some malicious. NIDS was capable of identifying malicious traffic effectively where antivirus created false negatives. ML classifiers were used as the most efficient and traditional antivirus in that technique. The ML model was integrated into traditional IDS for detecting advanced threats and decreasing false positives.

Communication systems had two major challenges: privacy of user-specific data and computer security. The increasing uses of Internet connected devices had a certain increased to more number of vulnerabilities, which integrated assault on devices. Ayyaz-Ul-Haq Qureshi et al. [49] proposed a novel random NN depending on IDS (RNN-IDS) for detecting malicious activity. The RNN architecture consists of one way to pass the signals or information where data or signal transfers from the input layer to the hidden layer. In that technique, performance was estimated by training dissimilar numbers of input and hidden layer neurons through a learning tariff on standard NSL-KDD datasets for binary classification. The gradient descent algorithm (GDA) was used for classifying the binary class of NSL-KDD datasets. A large number of input neurons were needed to improve detection accuracy. Moreover, the proposed method detects only a few attacks.

Thus, Sanchit Nayyar et al. [50] proposed an LSTM based ML approach for examining intrusions in the network. That proposed approach was tested on the CICIDS2017 data set with several attacks such as Web based Brute Force, DoS Hulk, DoS sloworis, DoS GoldenEye DoS slowhttptest, DDoS LOIT and Patator based attacks. The neurons in the input and output layers were 77 and 2; 12 hidden layers were presented. That proposed approach requires an efficient algorithm for better classification. Nuno Oliveira et al. [51] suggested a sequential approach based on NIDS.

Moreover, the performance of multilayer perception (MLP), random forest and long short term memory was estimated on the CIDDS-001 dataset. That proposed model was evaluated on single flow and multi-flow. Above mentioned methods and their advantages are illustrated in Table IV.

TABLE IV. NETWORK BASED IDS

Author	Method	Datasets	Classifier	Outcomes	Advantages	Disadvantages
Basant Subba in 2016 [41]	ANN	NSL-KDD	ANN	A-95.05 % DR-95.05 %	High accuracy and detection rate	More processing time for large neural network
Norbert Adam in 2017 [42]	NNIDS	Train.txt	ANN	CR-1 LR-0.5	Recognize learned malicious activities	Spanning port per switch was allowed
Vrushali D.Mane in 2018 [43]	BPANN	KDD 99	ANN	A-98.0% DR-92.80 %	Minimize FAR	Sensitive to noise data
Hossein Gharaee in 2016 [44]	GA and SVM	KDD CUP 99 and UNSW-NB15	SVM	A-99.05 % TP-98.47 % FP-0.4%	High accuracy and low FPR	SVM was not suitable for large datasets
Kai Zhang in 2019 [45]	IACF	LLDOS 1.0	SPA	A-90%	Efficient in alert correlation	Less information in the infiltration scenario
Basant Subba in 2016 [46]	GTBFA	DARPA and IITG Lab	SVS	A-98.55 % DR-91.87 %	High accuracy	The detection rate was low against non-critical vulnerabilities
Priyadarsi Nanda in 2016 [47]	LSSVM	KDD Cup 99, NSL-KDD and Kyoto 2006+	SVM	A-99.94 % DR-98.93 % FR-0.28%	Better accuracy and low computational cost	Not suitable for unbalanced sample distribution condition
Sanjay Kumar in 2016 [48]	ML	KDD99, DARPA 1998/1999 and ISCX 2012	ML classifier	A-99.4% DR-82% TP-99.6% FP-1.8%	Detect threats with high accuracy	Classifier modelled using few malware functions
Ayyaz-Ul-Haq Qureshi in 2018 [49]	RNN-IDS	NSL-KDD	GDA	A-94.50 % P-98.9% D-95.3% TP-95.31 % FP-1.28%	High precision value	More number of input and hidden neurons are required for better efficiency
Sanchit Nayyar et al. [50]	RNN-LSTM	CICIDS 2017 data set	RNN	A-96%	More attacks are founded	Need to develop an effective classification algorithm
Nuno Oliveira et al. [51]	Sequential approach	CIDDS-001 dataset	RF, MLP and LSTM	A-99.94 % F-91.66 %	Anomaly detection improved by sequential flow	Need to improve network optimization

2) Host-based IDS (HIDS): Host-based IDS protects our systems from many harmful intrusions occurring in the network. The host-based approach detects normal behavior by sequencing the system call. System sequences in the network create the sequences.

Chawla et al. [52] proposed Convolutional Neural Network (CNN) - Gated Recurrent Unit (GRU) language model for the just released Australian Defense Force Academy Linux Dataset (ADFA-LD). Training time is reduced by implementing CRR in the place of normal LSTM. Normal call sequence training is given to the model. In that approach, the next integer was predicted through training provided probability distribution. That proposed method had a lower convergence speed. Hence, Robin Gassais et al. [53] proposed host-based automated IDS by combining machine learning (ML) algorithms and tracing techniques to improve the convergence speed. That proposed approach used Random Forest (FT) Gradient Boosted Trees (GBT) for intrusion detection. Information on tracing was obtained from user space and kernel space. Adaptation based tuning for new devices are explained in that approach. The system has shown better accuracy in detecting threats and alerting the system. In that proposed approach, the network was analyzed only in series. Thus, it does not provide higher detection accuracy.

By focusing on higher network accuracy, Prachi Deshpande et al. [54] proposed HBIDS by analyzing system call traces and alerting users if any threats were detected. K-nearest neighbor (kNN) was used as a classifier for tracing. The kNN allows easy inclusion of new training data. The network was analyzed through the frequency of failed system calls over successful system calls. The host-based IDS methods and its outcome are illustrated in Table V.

TABLE V. HOST BASED IDS METHODS

Author	Method	Dataset	Outcome	Advantages	Drawbacks
Chawla et al. [52]	Convolutional Neural Network Gated Recurrent Unit	Australian Defence Force Academy Linux Dataset	TDR - 100% FAR - 60%.	More accurate with reduced training time.	Slow convergence speed
Robin Gassais et al. [53]	Random Forest Gradient Boosted Trees	User Space Kernel Space	DT-1.23 RF-1.68 GBT 9.28 SVM 6.79 MLP 2.16	Improvement in detection time.	Analyzing only in series, not parallel and need learning for detection
Prachi Deshpande et al. [54]	K-nearest neighbour Classifier	-	A- 96%	Improvement in Accuracy	Delayed Detection time

C. Network Intrusion Detection Approaches

Since there are several methods were suggested by the authors to deal with intrusions into the network. Previous sections analyze host-based, specification-based, network-based and anomaly-based methods. Those methods do not perform well in detecting all kinds of attacks and have reduced detection accuracy. Moreover, the network intrusions are detected by clustering, hybrid and evolutionary algorithms reviewed in the upcoming section.

1) Clustering based NIDS: The clustering-based approaches are aimed at reducing network complexities and easing detection accuracy. Some of the clustering-based NIDS are listed below.

Luiz Fernando Carvalho et al. [55] proposed an unsupervised learning approach for NIDS for extracting features. A modified ant colony optimization algorithm was used to optimize the multi-dimensional flow of the network. That proposed method offers lower detection accuracy; hence, to overcome these drawbacks, Yanqing Yang et al. [56] proposed a fuzzy aggregation method using a modified density peak clustering algorithm (MDPCA) and deep belief networks (DBNs). In that proposed approach, the training sets were divided into sub-sets with the same number of attributes of various cluster centers. That proposed method enhances automatic feature selection.

Moreover, Wei Liang et al. [57] proposed a multi-feature data clustering optimization model to improve impersonation attacks' detection accuracy. Security coefficients and weighted distances were categorized based on priority thresholds in that proposed method. The similarity of multi-feature data was examined through distance metrics. That proposed algorithm halted the clustering while reaching the preset iteration or obtaining the best cluster.

2) Evolutionary algorithm: Evolutionary computing methods are also known as bio-inspired algorithms that are processed based on the behavior of biological organisms. Bio-inspired algorithms are widely applied in many fields due to their simple processing. Evolutionary algorithm-based NIDS is given below.

Vajiheh Hajisalem and Shahram Babaie [58] proposed hybrid Artificial Bee Colony (ABC) and Artificial Fish Swarm (AFS) algorithms. In addition, the fuzzy C-means clustering and Correlation-based Feature Selection (CFS) techniques were adopted to avoid unnecessary features. The CART technique formulated the fuzzy rules for classifying normal and intrusion data in that approach. Further improving network intrusion classification, Chaouki Khammassi and Saoussen Krichen [59] proposed a genetic algorithm (GA) and logistic regression based learning algorithm. That proposed method was evaluated under the KDD99 dataset and the UNSW-NB15 dataset. In addition, three decision tree classifiers were adopted for performance evaluation. The performance measures examined for the KDD99 dataset are illustrated in the table.

3) Classification based NIDS: Classification-based approaches are adopted for classifying malicious and normal behavior of the network. The classifier is developed by training data and classifies network behavior in classification-based methods. The classification approaches are carried out in either a multi-class or single-class manner. Some of the classification approaches for NIDS are given below.

Jiyeon Kim et al. [60] proposed a CNN-based NIDS to examine the DOS attack evaluated on the KDD CUP 1999 dataset comprised of DOS, U2R, R2L and probing attacks. That proposed model detects attacks belonging to a similar category. Moreover, the CSE-CIC-IDS2018 is also used for examining advanced IDS attacks. That proposed method offers lower detection accuracy of 93%. To enhance detection accuracy for examining cyber-attacks, Guo Pu et al. [61] proposed a hybrid unsupervised cluster based NIDS by combining OC-SVM and subspace clustering methods. This proposed method was validated under the NSL-KDD dataset. The OCSVM was suitable for unlabelled data in which the data had a normal class only. That maps both the feature space with data under the kernel. The feature selection must be improved further for better performance.

Yang Jia et al. [62] proposed a new deep NN (NDNN) model for NIDS that comprised four hidden layers evaluated on KDD99 and NSL-KDD. In that model, the input, hidden, and output layers consist of 41 neurons, 100 neurons, and 5 neurons, respectively. The proposed model performs better in KDD99 datasets but cannot provide better detection accuracy. But the KDD99 provides better performance in different kinds of attacks. Ahmed Iqbal and Shabib Aftab [63] proposed a feed-forward and pattern recognition (PR) ANN model to improve detection accuracy. The ANN was trained by scaled conjugate gradient, training functions and Bayesian regularization. That feed-forward network had multilayer neurons that were trained by Bayesian regularization. Then the PR was trained by a scaled conjugate gradient function. Arun Nagaraja et al. [64] proposed a UTTAMA classifier for detecting network intrusion; moreover, the Apriori algorithm based Frequent Pattern (FP) max algorithm was used. That

proposed classifier was tested under KDD-41 and KDD-19 datasets.

4) Hybrid approaches: Hybrid approaches are developed by adding various classification algorithms, methods and techniques. Some of the hybrid approaches are listed below.

Ansam Khraisat et al. [65] proposed a hybrid C5 decision tree classifier and OC-SVM, estimated on network security laboratory-knowledge discovery in a database (NSL-KDD). The proposed model used the attributes of benign samples, which does not use the data from other samples. The OCSVM classifier converts instances into a high dimensional attributes space and locates a suitable boundary hyperplane. That approach can detect the intrusions with only limited samples whereas, it has lower accuracy. Muhammad Ashfaq Khan [66] proposed a hybrid convolutional recurrent neural network-based NIDS. That method was estimated on the CSE-CIC-DS2018 dataset that comprised seven types of attacks. Moreover, the network was mixed with traffic and non-traffic nodes to examine the suggested method. That suggested method was only investigated on the single dataset.

Different NIDS approaches are illustrated in Table VI.

TABLE VI. DIFFERENT APPROACHES FOR NETWORK INTRUSION DETECTION

Author	Method	Performance measures	Advantages	Drawbacks
Luiz Fernando Carvalho et al. [55]	Ant colony optimized digital signature	TPR is 93% for 1% FPR	Suitable for anomaly detection in large scale network	Lower detection accuracy
Yanqing Yang et al. [56]	MDPCA and DBN	A-90.21% DR-96.22% FPR-17.15%	Reduces the complexities in training sub-sets	Cannot detect the R2L and U2L attacks
Wei Liang et al. [57]	Multi-feature data clustering	DA-0.95 for 75 features	Detection accuracy is improved during a high overlap	Only detect the known attacks
Vajiheh Hajisalem and Shahram Babaie [58]	CFS techniques	FPR-0.01% DR-99%	Lower overhead	Need to reduce network complexity
Chaouki Khammassi and Saoussen Krichen [59]	genetic algorithm (GA) and logistic regression based learning algorithm	FAR-99.81% DR-1.105%	Better accuracy in detecting DOS attack	Need to reduce the misclassified results
Jiyeon Kim et al. [60]	CNN	P-1 A-93% R-1 F-1	New attacks are identified through the CSE-CIC-IDS2018 dataset	Only detect the attack belongs to the same category
Guo Pu et al. [61]	SSC-OCSVM	DR-1 @ 0.05 false alarm rate	Detect large fraction of attacks in lower FAR	Lower detection accuracy
Yang Jia et al. [62]	NDNN	A-0.999 F-0.9998 R-0.9997	KDD99 dataset provides the best performance in all kinds of attacks detection	More network simulation experiments are needed
Ahmed Iqbal and Shabib Aftab [63]	FFANN	A-99.8356% MSE-0.0050	KDD CUP 99 examines different attacks	More improvements needed
Arun Nagaraja et al. [64]	UTTAMA	A-99.952%	A new membership function achieves a dimensional reduction	New distance functions are needed to improve performance efficacy.
Ansam Khraisat et al. [65]	OCSVM	A-83.24%	Detect the intrusions with fewer samples	Lower accuracy for detecting different tasks
Muhammad Ashfaq Khan [66]	HCRNNIDS	A-97.75% P-0.9633 R-0.9712 F-0.976 DR-0.97 FAR-2.5	Provides sufficient security against malicious nodes	The proposed method is tested on a single dataset only

5) Discussion on traditional methods

- The anomaly-based methods follow some detection rules based on some considerations thus, it is not capable of detecting all faults in the network. Moreover, the rule-defining process is degraded by the protocols, whereas the custom protocols also affect the rule-defining process.
- The clustering-based NIDS methods reduce the computational complexities in the network by grouping large datasets. In contrast, these methods affect detection rates in positive and negative ways. Moreover, the clustering approaches consume more time and can be affected by local minima.
- Classification-based NIDS approaches to improve the detection accuracy of network intrusion. Due to its adaptive nature, more data can be trained and tested using the ANN. On the other hand, the detection performance is improved by adding more layers within the network. Thus resource utilization is increased, resulting in over-fitting issues. Furthermore, that approaches needs relevant information for detecting unknown attacks.
- In evolutionary algorithm based NIDS, the detection accuracy and intrusion detection performance depend

on the algorithm used. The meta-heuristic algorithms adopt the behavior of living organism; thus, it does not require prior information about the network. The evolutionary algorithms do not affect by network noises. At the same time, the evolutionary algorithms are hard to map with network related issues.

- Compared with other methods, hybrid NIDS approaches have better performance but require high resources due to the combination of multiple techniques.

From the above analysis, it was noted that the rule-based approaches are affected by the protocols and clustering approaches increases the complexity due to large datasets and the classification based approaches needs more layers to provide an accurate results.

D. Deep Learning (DL) and Machine Learning based NIDS

Several methods are reviewed in the literature for detecting intrusion in the network, but they lack data management and feature learning. Both are considered important issues in IDS. Thus the DL based approaches are developed for detecting security threats in networks. Some of the DL based approaches are listed below.

Quamar Niyaz et al. [67] proposed a DL approach for detecting intrusion in the network in which self-taught learning (STL) was adopted. That proposed STL method comprised two stages wherein the representation of good features was learnt from unbalanced data collection. Then that learn representation was employed for labelled data that was adopted for classification. In that approach, the NSL-KDD dataset was adopted, which was minimized version of the KDD Cup 99 dataset. The sparse auto-encoder and soft-max regression matrix-based IDS were implemented in that paper. Furthermore, the performance of the IDS was enhanced by adding an extension of auto-encoders and stacked auto-encoder.

Hongpo Zhang et al. [68] proposed a denoising auto-encoder (DAE) and a weight loss function to select appropriate features for detecting intrusion by minimizing the feature dimensionality. Then, the chosen features were categorized through the multilayer perceptron (MLP) to identify the intrusion. In that proposed method, one key feature selection adds the weights to several samples' loss functions, which eases the feature selection. In that method, 12 features were chosen among 202 features with a 5.9% selection ratio. In MLP, two hidden layers were used for classification, which yielded an accuracy of 98.80%.

Fahimeh Farahnakian and Jukka Heikkonen [69] proposed a deep auto-encoder (DAE)-based method to detect network attacks caused by several vulnerabilities. That proposed DAE was trained in a greedy layer-wise fashion to evade the over-fitting and local optima. That proposed approach was executed through training and testing phases. During the training process, the system did the adoption of the training dataset and the creation of the DAE model. During the testing phase, a model was used for labelling test data. The input layer denotes 117 features of the dataset, and then the hidden layer chooses 32 features. In that method, the sigmoid function was chosen for hidden layers.

Vinayakumar et al. [70] proposed a deep NN (DNN) for detecting intrusion and classifying cyber-attacks. The Apache Spark cluster computing platform designed the scalable computing architecture in that proposed method. That proposed architecture explores the processing competence of a general purpose graphical processing unit (GPGPU) for rapid network investigation. The proposed architecture of DNN comprises five hidden layers for extracting complex features and pattern recognition ability. That proposed method was executed for different classifiers such as KDDCup 99, NSL-KDD, UNSW NB-15 and WSN-DS. Moreover, the result implies that the dataset of KDDCup99 and NSLKDD offer higher accuracy in the 95% to 99% range. The performance measures for binary classification of the KDDCup99 dataset with five layers of DNN are illustrated in the table.

Faten Louati and Farah Barika [71] proposed a DL based multi-agent system (MAS) for detection by combining the multi-agent features with DL algorithms. In that approach, the agents were generated by three algorithms: MLP, auto-encoder and k-nearest neighbour in which the autoencoder was used for feature reduction. In contrast, MLP and autoencoder were used for classification. In that approach, the MLP was a subset of DNN that was based on the backpropagation (BP) algorithm. Moreover, in autoencoders, the output layer had a similar number of nodes as the input layer. The auto encoders minimize the number of features from 120 to 10. In addition, the MAS enhances the network performance through proactivity and reactivity, which eases intrusion detection. The MAS comprises a pre-processor agent, reducer agent, classifier agent and decision-maker agent.

Soosan Naderi Mighan and Mohsen Kahani [72] proposed a hybrid DL and support vector machine (SVM) for feature extraction and classification. The stacked autoencoder (SAE) was used to reduce feature sets' dimensional reduction. In addition, the SVM was adopted for classification. In the pre-processing stage, symbolic features were converted into numeric values that range from zero to several symbols. The data normalization step was executed to minimize the dimension for all attributes. Then the latent features were extracted in the second phase; after that, the DL approach detected the attack. That suggested model was investigated on the ISCX IDS UNB dataset and it results in faster execution time.

Kasongo and Yanxia Sun [73] proposed feed-forward DNN (FFDNN) along with wrapper based feature selection (WFEU). The proposed method used an extra tree algorithm for optimum feature selection. That proposed method was evaluated under AWID and UNSW-NB15 datasets. The UNSW-NB15 dataset comprised 39 numerical features, three input nominal features had the feature type of binary, float and integer. In the initial phase, the FFDNN was implemented; after that, the WFEU was included. The experimental results demonstrate that the wrapper-based feature extraction was effective for the UNSW-NB15 dataset. In AWID, the

proposed model was implemented with a whole set of features; after that, 26 features were used from 154 attributes.

Sandeep Gurung et al. [74] proposed a DL and sparse auto-encoders for feature learning. That proposed method was trained by the NSL-KDD dataset that yields the output value as 0 (normal user) or 1 (intruder). In a pre-processing step, the numeric parameters were replaced instead of non-numeric parameters then the data was normalized. That proposed method reduced the false alarm rate lower than the signature-based method. Mike Nkongolo et al. [75] proposed a novel dataset, UGRansome1819, to detect unknown network attacks like zero-day threats. That proposed dataset benefited from unknown attacks that were not explored before and could not be observed by known attacks that were more efficient than the KDD99 and NSL-KDD datasets.

By analyzing several DL-based intrusion detection approaches, it has been concluded that deep network models examine all types of intrusion in the network. Furthermore, the deep networks can classify the intrusion type thus, it can be useful for detecting unknown attacks in the network. The node with an Internet protocol (IP) address and Internet produces network traffic that blocks the service to users [76]. However, the ML based methods are applicable in ad-click prognosis systems, and a two-way authentication system provides a secure connection between digital environments [77, 78]. If the attack arises from several distributed hosts, it will cause distributed denial of service attack (DDoS) that is more harmful than normal DoS [79, 80]. Illegitimate users create the DDoS to deny the server provided services [81]. Moreover, the system performance is improved by adding auto-encoders that increase detection accuracy.

The DL-based IDS with its pros and cons are illustrated in Table VII.

TABLE VII. DL BASED METHODS FOR IDS

Author	Method	Datasets	Classifier	Outcome	Advantages	Disadvantages
Quamar Niyaz et al. in 2016 [67]	DL	NSL-KDD dataset	Auto encoder	-	Improved performance by adding NB-tree and random-tree classifiers	The proposed method was not evaluated under real time IDS
Hongpo Zhang et al. in 2018 [68]	denoising auto-encoder (DAE)	UNSW-NB dataset	MLP	A-98.80% F-0.952 P-95.98% R-94.43% FPR-0.57%	Suitable for high speed network	-
Fahimeh Farahnakian and Jukka Heikkonen in 2018 [69]	DAE	KDD-CUP 99	Softmax	A-94.71% FA-0.42% DR-95.53%	A greedy unsupervised layer-wise training mechanism improves performance	The imposition of sparsity constraints in autoencoders is not illustrated in the paper.
Vinayakumar et al. in 2019 [70]	DNN	KDD-CUP 99, NSL-KDD and UNSW NB-15, WSN-DS	DT, AB, RF, LR, NB, KNN, SVM-rbf	A-0.927 P-0.994 R-0.915 F-0.953 (KDDCup99)	Better performance than machine learning classifiers	Failed to monitor DNS and BGP
Faten Louati and Farah Barika in 2020 [71]	DL-MAS	KDDCup99		A-99.95% P-00.32% FPR-0.17% FNR-0.68% TNR-99.83% TPR-99.32% DR-99.32%	The system tolerates the failure of one or more agents	Lower mobility and cloning
Soosan Naderi Mighan and Mohsen Kahani in 2018 [72]	DL-SAE	ISCX IDS UNB dataset	DL-SVM	A-0.902 TPR-0.903 TNR-0.902 P-0.903 R, F-0.903	Faster execution time	Not capable of managing large scale network traffic
Kasongo and Yanxia Sun in 2020 [73]	FFDNN	AWID, UNSW-NB15	SVM, decision tree (DT), random forest (RF), naïve Bayes (NB) and k-nearest neighbour (kNN)	A-99.66%	Adapted for wireless application	Lower testing accuracy, 87.48% for binary classification
Sandeep Gurung et al. [74]	Sparse AE	NSL-KDD dataset	Logistic classifier	A-87.2% P-84.7% R-92.8% Sp-80.7% NPV-90.7%	Used in any server which monitors the network activity	Lower accuracy

1) Discussion on DL and ML approaches: As mentioned earlier, the traditional methods have some limitations in intrusion detection; thus, the DL and ML based methods are used to resolve such issues. The insight on DL and ML based approaches are discussed below:

- The ML based methods provide better results in higher detection accuracy, but their performance completely relies on the data training.
- The DL methods efficiently examine enormous data in the network, but optimized layers improve their performance. Thus if the DL is used it must be used with the optimization algorithm to improve the computational efficacy.
- In a supervised learning based ML approach, classification is an important task; however, manual data labelling consumes more time for processing. Thus, the data labeling must be improved to get the benefits in ML approaches.
- The methods reviewed in the above section have used commonly used datasets such as KDD, CUP 99 etc., which do not contains the updated data of new types of attacks. While applying the methods to these datasets, the method's robustness cannot be determined. On the other hand, constructing a new dataset is expensive and needs expert knowledge.
- The better performance of IDS is not guaranteed while using ML with the dataset without real-world samples.

Findings: From these analysis, it is found that the method should be improved in terms of data collection, feature extraction and classification. Moreover, the inclusion of up-to-date dataset is also important factor to examine the unknown attacks. Moreover, these findings are more sufficient for proving efficient solution to improve the IDS against various attacks.

E. Limitations of NIDS

Although several methods are available for network intrusion detection, some limitations are listed below.

- Despite the ongoing research on IDS, intrusion detection techniques still have some drawbacks, such as slow detection time and high false detection. Due to the high detection time, it is hard to make working these methodologies for high speed networks.
- It is impossible to run soft-computing techniques on huge data with many features and imbalanced data. For analyzing such huge data, efficient sampling and feature selection techniques must be used.
- Parallelly coordinated IDS are needed for large-scale and fast computing networks [82].
- Recently developed approaches perform better in detecting network intrusions but do not detect all types of attacks. Moreover, the available datasets comprised only a few attacks; thus, it is not effective for detecting all types of attacks. The up-to-date datasets may reveal new attacks in the network; thus, most of the recent IDS cannot detect several types of attacks due to the unavailability of up-to-date datasets.
- The important issue with anomaly-based method is that it can examine the zero-day attack when properly modelled. The improper modelling of anomaly-based methods will raise the false alarm rate.
- The dataset must be integrated with the DL models to identify more attacks in the network to learn many patterns. In these cases, the dataset generation will be expensive.
- Most importantly, the unbalancing dataset may degrade network performance by reducing detection accuracy; thus, a balanced dataset is required for better dataset performance.
- The major challenge in IDS is the execution in the real-world environment, whereas most existing methods are not validated using a public dataset in the lab.
- Several methods suggested in existing works are complex in structure; this may cause extra overhead for the network process.
- To minimize the routing overhead of the network by proper feature selection, the researchers propose optimization algorithms, but the convergence speed of the existing algorithms will affects the feature extraction.
- The ML approach has some limitations, such as handling raw, high dimensional data and unlabeled data. Thus it cannot provide better classification in the presence of large datasets and complex data labelling. Thus the ML approaches are not suitable in the case of multi-classification function.
- The IDS are also adopted in IoT applications thus that deal with several sensor nodes in such methods. Not all kinds of IDS are suitable for this sense; only lightweight IDS is preferred in these applications due to minimum power utilization.
- IDS plays a major role in providing better system security than other systems. However, new malicious attacks are occurring in large amounts, making providing system security within computer networks a tedious task. Therefore, systematically updating available datasets would be the need of the hour.
- The challenges associated with IDS are false alarm rate, low detection rate, unbalanced datasets and response time. Misuse or signature based IDS are usually accompanied by some degree of false-positive alarm rates and are inefficient in detecting unknown or novel attacks. The main challenge of this system is updating the signatures of harmful intrusions. The challenges associated with anomaly based IDS are miscalculation in detection, lack of speed, difficulty in alerting and unbalanced datasheets.

V. CONCLUSION

The IDS is a sufficient mechanism for securing the network against intrusions; thus, a comprehensive review of this topic is needed. This survey paper aimed to better understand intrusion detection in the network in accordance with different aspects. Different methods reviewed in this paper provide perceptive growth in intrusion detection. A wide range of methods was developed on this topic, each with advantages and drawbacks. A comparative analysis of different methods is given in this paper. Initially, several attacks occur in the networks were analyzed. After that, the NIDS is categorized under the detection technique and monitor platform. Further, the detection techniques are specification-based, anomaly-based and hybrid methods. Similarly, depending on the monitoring platform, the monitoring platform is classified as network-based or host-based. Additionally, classification-based, algorithm-based, clustering-based IDS are reviewed in detail. Tables provided deep insight into these methods and their pros and cons in each section. Different datasets are used in several methods to show the performance efficacy of the suggested method. It is found that most existing approaches detect few attacks only due to the availability of up-to-date datasets. By analyzing different methods, IDS is limited; the signature-based IDS are not effective for detecting the type of attack; thus, efficient IDS are modelled for detecting several kinds of attacks on the network.

The anomaly-based IDS effectively detect different types of attacks in the network, resulting in a higher false alarm rate; hence in future work, the anomaly-based approach will be improved by minimizing the false alarm rate. Even though the anomaly-based approaches can detect the zero-day attack, they cannot provide the desired outcome if it is not properly designed. To provide these issues, the ML and DL approaches are developed. The important task in intrusion detection is feature extraction. In the view of feature extraction, the DL outperforms ML by automatically extracting features. In addition, the survey shows that the classification-based methods had better performance in intrusion detection than all other methods. In addition, DL models are more effective for classification. Thus, it is concluded that DL approaches are more effective for detecting unknown attacks in the network. Most DL-based approaches are validated through different datasets in which the detection accuracy is based on feature selection. For this purpose, optimization algorithms are adopted. But single algorithms are not improving optimum feature selection. Since no methods were developed for unknown attack detection in the network with optimized DL. On the other hand, it is observed that the lack of an up-to-date dataset degrades intrusion detection in the network. This review cannot affirm the best method to detect unknown attacks in the network, but it suggests an effective way to detect unknown attacks. The future research direction of unknown attack detection is deliberated in the next section.

A. Future Works

The future scope of IDS is listed below.

This review verified that DL approaches are more effective, but their performance can be improved by adding an efficient algorithm. Hence, hybrid DL approaches are suggested for detecting known and unknown attacks in future work. In order to improve the detection of unknown attacks, it is important to use the up-to-date method. Future work will focus on unknown attack detection in the network; hence, the recently developed UGRansome1819 dataset is suggested. The intrusion detection not only improved by the dataset but also improved by adding an efficient feature section approach. In fact, hybrid pattern search whale optimization algorithm will be effective for optimal feature selection. Pattern search is a non-derivative algorithm that is suitable for updating an optimum weight as well as the whale optimization algorithm (WOA) has a better convergence speed. Hence adopting both these algorithms will improve the feature selection. The important task after feature selection is classification, which provides the final outcome about the attacks in the network. At the same time, the lack of processing speed will reduce the detection accuracy. Furthermore, the hybrid bi-directional long short term memory (Bi-LSTM) with the gated recurrent unit (GRU) has to be implemented to classify unknown attacks in the network. These improved feature extraction and classification approaches will be effective solutions for detecting unknown attacks in the network.

REFERENCES

[1] M. Ozkan-Okay, R. Samet, Ö. Aslan, & D. Gupta, "A Comprehensive Systematic Literature Review on Intrusion Detection Systems," IEEE Access, 2021.

[2] R.V. Mendonça, A.A. Teodoro, R.L. Rosa, M. Saadi, D.C. Melgarejo, P.H. Nardelli, & D.Z. Rodríguez, "Intrusion detection system based on fast hierarchical deep convolutional neural network," IEEE Access, vol. 9, pp. 61024-61034, 2021.

[3] M. Al-Qatf, Y. Lasheng, M. Al-Habib, & K. Al-Sabahi, "Deep learning approach combining sparse autoencoder with SVM for network intrusion detection," Ieee Access, vol. 6, pp. 52843-52856, 2018.

[4] A. Borkar, A. Donode, & A. Kumari, "A survey on Intrusion Detection System (IDS) and Internal Intrusion Detection and protection system (IIDPS)," In 2017 International conference on inventive computing and informatics (ICICI), IEEE, pp. 949-953, 2017.

[5] W. Li, S. Tug, W. Meng, & Y. Wang, "Designing collaborative blockchained signature-based intrusion detection in IoT environments," Future Generation Computer Systems, vol. 96, pp. 481-489, 2019.

[6] M. Torabi, N.I. Udzir, M.T. Abdullah, & R. Yaakob, "A review on feature selection and ensemble techniques for intrusion detection system," network, vol. 1, pp. 2, 2021.

[7] M. Uğurlu, and İ.A. Doğru, "A survey on deep learning based intrusion detection system," In 2019 4th International Conference on Computer Science and Engineering (UBMK), IEEE pp. 223-228, 2019.

[8] M. Almseidin, M. Alzubi, S. Kovacs, & M. Alkasassbeh, "Evaluation of machine learning algorithms for intrusion detection system," In 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), IEEE, pp. 000277-000282, 2017.

[9] M.Y. AlYousef, & N.T. Abdelmajeed, "Dynamically Detecting Security Threats and Updating a Signature-Based Intrusion Detection System's Database," Procedia Computer Science, vol. 159, pp. 1507-1516, 2019.

[10] M. Dua, "Machine learning approach to IDS: A comprehensive review," In 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE pp. 117-121, 2019.

[11] M. Maithem, and G.A. Al-sultany, "Network intrusion detection system using deep neural networks," In Journal of Physics: Conference Series, IOP Publishing vol. 1804, no. 1, pp. 012138, 2021.

[12] G. Karatas, O. Demir, and O.K. Sahingoz, "Deep learning in intrusion detection systems, In 2018 International Congress on Big Data," Deep

Learning and Fighting Cyber Terrorism (IBIGDELFT), IEEE pp. 113-116, 2018.

[13] O.M.A. Alsyaibani, E. Utami, and A.D. Hartanto, "Survey on Deep Learning Based Intrusion Detection System," Telematika, vol. 14, no. 2, pp. 86-100, 2021.

[14] A.R. Khan, M. Kashif, R.H. Jhaveri, R. Raut, T. Saba, and S.A. Bahaj, "Deep learning for intrusion detection and security of Internet of things (IoT): current analysis, challenges, and possible solutions," Security and Communication Networks, vol. 2022, 2022.

[15] Y. Li, R. Liu, X. Liu, H. Li, & Q. Sun, "Research on information security risk analysis and prevention technology of network communication based on cloud computing algorithm," In Journal of Physics: Conference Series, IOP Publishing vol. 1982, no. 1, pp. 012129, 2021.

[16] T.H. Hadi, "Types of Attacks in Wireless Communication Networks," Webology, vol. 19, no. 1, 2022.

[17] M.V Pawar, & J. Anuradha, "Network security and types of attacks in network," Procedia Computer Science, vol. 48, pp. 503-506, 2015.

[18] P. Amish, & V.B. Vaghela, "Detection and prevention of wormhole attack in wireless sensor network using AOMDV protocol," Procedia computer science, vol. 79, pp. 700-707, 2016.

[19] A.U. Rehman, S.U. Rehman, & H. Raheem, "Sinkhole attacks in wireless sensor networks: a survey," Wireless Personal Communications, vol. 106, no. 4, pp. 2291-2313, 2019.

[20] Z. Gavric, & D. Simic, "Overview of DOS attacks on wireless sensor networks and experimental results for simulation of interference attacks," Ingeniería e Investigación, vol. 38, no. 1, pp. 130-138, 2018.

[21] B. Bhushan, & G. Sahoo, "Recent advances in attacks, technical challenges, vulnerabilities and their countermeasures in wireless sensor networks," Wireless Personal Communications, vol. 98, no. 2, pp. 2037-2077, 2018.

[22] M. Conti, N. Dragoni, and V. Lesyk, "A survey of man in the middle attacks," IEEE communications surveys & tutorials, vol. 18, no. 3, pp. 2027-2051, 2016.

[23] K.L. Chiew, K.S.C. Yong, and C.L. Tan, "A survey of phishing attacks: Their types, vectors and technical approaches," Expert Systems with Applications, vol. 106, pp. 1-20, 2018.

[24] A. Thakkar, & R. Lohiya, "A survey on intrusion detection system: feature selection, model, performance measures, application perspective, challenges, and future research directions," Artificial Intelligence Review, pp. 1-111, 2021.

[25] A. Khraisat, I. Gondal, P. Vamplew, & J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," Cybersecurity, vol. 2, no. 1, pp. 1-22, 2019.

[26] M. Almansor, & K.B. Gan, "Intrusion detection systems: principles and perspectives," Journal of Multidisciplinary Engineering Science Studies, vol. 4, no. 11, pp. 2458-2925, 2018.

[27] A. Adnan, A. Muhammed, A.A Abd Ghani, A. Abdullah, & F. Hakim, "An Intrusion Detection System for the Internet of Things Based on Machine Learning: Review and Challenges," Symmetry, vol. 13, no. 6, pp. 1011, 2021.

[28] M. Masdari, and H. Khezri, "A survey and taxonomy of the fuzzy signature-based intrusion detection systems," Applied Soft Computing, vol. 92, pp. 106301, 2020.

[29] A. Lazarevic, V. Kumar, & J. Srivastava, "Intrusion detection: A survey," In Managing cyber threats, Springer, Boston, MA pp. 19-78, 2005.

[30] R.A. Beyah, M.C. Holloway, & J.A. Copeland, "Invisible trojan: an architecture, implementation and detection method," In The 2002 45th Midwest Symposium on Circuits and Systems, 2002. MWSCAS-2002., IEEE, vol. 3, pp. III-III, 2002.

[31] S. Zavrak, & M. İskefiyeli, "Anomaly-based intrusion detection from network flow features using variational autoencoder," IEEE Access, vol. 8, pp. 108346-108358, 2020.

[32] S. Dwivedi, M. Vardhan, S. Tripathi, & A.K. Shukla, "Implementation of adaptive scheme in evolutionary technique for anomaly-based intrusion detection," Evolutionary Intelligence, vol. 13, no. 1, pp. 103-117, 2020.

[33] B.A. Tama, M. Comuzzi, & K.H. Rhee, "TSE-IDS: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system," IEEE Access, vol. 7, no. 94497-94507, 2019.

[34] N.T. Van, & T.N. Thinh, "An anomaly-based network intrusion detection system using deep learning," In 2017 international conference on system science and engineering (ICSSE), IEEE, pp. 210-214, 2017.

[35] R. Kumar, & D. Sharma, "HyINT: signature-anomaly intrusion detection system, In 2018 9th International Conference on Computing," Communication and Networking Technologies (ICCCNT), IEEE, pp. 1-7, 2018.

[36] A. Le, J. Loo, K.K. Chai, & M. Aiash, "A specification-based IDS for detecting attacks on RPL-based network topology," information, vol. 7, no. 2, pp. 25, 2016.

[37] H. Esquivel-Vargas, M. Caselli, & A. Peter, "Automatic deployment of specification-based intrusion detection in the BACnet protocol," In Proceedings of the 2017 Workshop on Cyber-Physical Systems Security and PrivaCy, pp. 25-36, 2017.

[38] A. Althubaity, H. Ji, T. Gong, M. Nixon, R. Ammar, & S. Han, "ARM: A hybrid specification-based intrusion detection system for rank attacks in 6TiSCH networks," In 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), IEEE, pp. 1-8, 2017.

[39] S. Einy, C. Oz, & Y.D. Navaei, "The anomaly-and signature-based IDS for network security using hybrid inference systems," Mathematical Problems in Engineering, 2021.

[40] M. Mehmood, T. Javed, J. Nebhen, S. Abbas, R. Abid, G.R. Bojja, & M. Rizwan, "A hybrid approach for network intrusion detection," CMC-Comput. Mater. Contin, vol. 70, pp. 91-107, 2022.

[41] B. Subba, S. Biswas, & S. Karmakar, "A neural network based system for intrusion detection and attack classification," In 2016 Twenty Second National Conference on Communication (NCC), IEEE, pp. 1-6, 2016.

[42] N. Ádám, B. Madoš, A. Baláž, & T. Pavlik, "Artificial neural network based IDS," In 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI), IEEE, pp. 000159-000164, 2017.

[43] V.D. Mane, & S. Pawar, "Anomaly based ids using backpropagation neural network," International Journal of Computer Applications, vol. 136, no. 10, pp. 29-34, 2016.

[44] H. Gharaee, & H. Hosseinvand, "A new feature selection IDS based on genetic algorithm and SVM," In 2016 8th International Symposium on Telecommunications (IST), IEEE, pp. 139-144, 2016.

[45] K. Zhang, F. Zhao, S. Luo, Y. Xin, & H. Zhu, "An intrusion action-based IDS alert correlation analysis and prediction framework," IEEE Access, vol. 7, pp. 150540-150551, 2019.

[46] B. Subba, S. Biswas, & S. Karmakar, "False alarm reduction in signature-based IDS: game theory approach," Security and Communication Networks, vol. 9, no. 18, pp. 4863-4881, 2016.

[47] M.A. Ambusaidi, X. He, P. Nanda, & Z. Tan, "Building an intrusion detection system using a filter-based feature selection algorithm," IEEE transactions on computers, vol. 65, no. 10, pp. 2986-2998, 2016.

[48] S. Kumar, A. Viinikainen, & T. Hamalainen, "Machine learning classification model for network based intrusion detection system," In 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST), IEEE, pp. 242-249, 2016.

[49] H. Larijani, J. Ahmad, & N. Mtetwa, "A novel random neural network based approach for intrusion detection systems," In 2018 10th Computer Science and Electronic Engineering (CEEC), IEEE, pp. 50-55, 2018.

[50] S. Nayyar, S. Arora, & M. Singh, "Recurrent neural network based intrusion detection system," In 2020 International Conference on Communication and Signal Processing (ICCSP), IEEE. pp. 0136-0140, 2020.

[51] N. Oliveira, I. Praça, E. Maia, & O. Sousa, "Intelligent cyber-attack detection and classification for network-based intrusion detection systems," Applied Sciences, vol. 11, no. 4, pp. 1674, 2021.

[52] A. Chawla, B. Lee, S. Fallon, & P. Jacob, "Host based intrusion detection system with combined CNN/RNN model," In Joint European

Conference on Machine Learning and Knowledge Discovery in Databases, pp. 149-158, 2018. Springer, Cham.

[53] R. Gassais, N. Ezzati-Jivan, J.M. Fernandez, D. Aloise, & M.R. Dagenais, "Multi-level host-based intrusion detection system for Internet of things," Journal of Cloud Computing, vol. 9, no. 1, pp. 1-16, 2020.

[54] P. Deshpande, S.C. Sharma, S.K. Peddoju, & S. Junaid, "HIDS: A host based intrusion detection system for cloud computing environment," International Journal of System Assurance Engineering and Management, vol. 9, no. 3, pp. 567-576, 2018.

[55] L.F. Carvalho, S. Barbon Jr, L. de Souza Mendes, & M.L. Proenca Jr, "Unsupervised learning clustering and self-organized agents applied to help network management," Expert Systems with Applications, vol. 54, pp. 29-47, 2016.

[56] Y. Yang, K. Zheng, C. Wu, X. Niu, & Y. Yang, "Building an effective intrusion detection system using the modified density peak clustering algorithm and deep belief networks," Applied Sciences, vol. 9, no. 2, pp. 238, 2019.

[57] W. Liang, K.C. Li, J. Long, X. Kui, & A.Y. Zomaya, "An industrial network intrusion detection algorithm based on multifeature data clustering optimization model," IEEE Transactions on Industrial Informatics, vol. 16, no. 3, pp. 2063-2071, 2019.

[58] V. Hajisalem, & S. Babaie, "A hybrid intrusion detection system based on ABC-AFS algorithm for misuse and anomaly detection," Computer Networks, vol. 136, pp. 37-50, 2018.

[59] C. Khammassi, & S. Krichen, "A GA-LR wrapper approach for feature selection in network intrusion detection," computers & security, vol. 70, pp. 255-277, 2017.

[60] J. Kim, J. Kim, H. Kim, M. Shim, & E. Choi, "CNN-based network intrusion detection against denial-of-service attacks," Electronics, vol. 9, no. 6, pp. 916, 2020.

[61] G. Pu, L. Wang, J. Shen, & F. Dong, "A hybrid unsupervised clustering-based anomaly detection method," Tsinghua Science and Technology, vol. 26, no. 2, pp. 146-153, 2020.

[62] Y. Jia, M. Wang, & Y. Wang, "Network intrusion detection algorithm based on deep neural network," IET Information Security, vol. 13, no. 1, pp. 48-53, 2019.

[63] A. Iqbal, & S. Aftab, "A Feed-Forward and Pattern Recognition ANN Model for Network Intrusion Detection," International Journal of Computer Network & Information Security, vol. 11, no. 4, 2019.

[64] A. Nagaraja, & B. Uma, "UTTAMA: An intrusion detection system based on feature clustering and feature transformation," Foundations of Science, vol. 25, no. 4, pp. 1049-1075, 2020.

[65] A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, & A. Alazab, "Hybrid intrusion detection system based on the stacking ensemble of c5 decision tree classifier and one class support vector machine," Electronics, vol. 9, no. 1, pp. 173, 2020.

[66] M.A. Khan, "HCRNNIDS: hybrid convolutional recurrent neural network-based network intrusion detection system," Processes, vol. 9, no. 5, pp. 834, 2021.

[67] A. Javaid, Q. Niyaz, W. Sun, & M. Alam, "A deep learning approach for network intrusion detection system," Eai Endorsed Transactions on Security and Safety, vol. 3, no. 9, pp. e2, 2016.

[68] H. Zhang, C.Q. Wu, S. Gao, Z. Wang, Y. Xu, & Y. Liu, "An effective deep learning based scheme for network intrusion detection, In 2018 24th International Conference on Pattern Recognition (ICPR)," IEEE, pp. 682-687, 2018.

[69] F. Farahnakian, & J. Heikkonen, "A deep auto-encoder based approach for intrusion detection system, In 2018 20th International Conference on Advanced Communication Technology (ICACT)," IEEE, pp. 178-183, 2018.

[70] R. Vinayakumar, M. Alazab, K.P. Soman, P. Poornachandran, A. Al-Nemrat, & S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," IEEE Access, vol. 7, pp. 41525-41550, 2019.

[71] F. Louati, & F.B. Ktata, "A deep learning-based multi-agent system for intrusion detection," SN Applied Sciences, vol. 2, no. 4, pp. 1-13, 2020.

[72] S.N. Mighan, & M. Kahani, "Deep learning based latent feature extraction for intrusion detection, In Electrical Engineering (ICEE)," Iranian Conference on, IEEE, pp. 1511-1516, 2018.

[73] S.M. Kasongo, & Y. Sun, "A deep learning method with wrapper based feature extraction for wireless intrusion detection system," Computers & Security, vol. 92, pp. 101752, 2020.

[74] S. Gurung, M.K. Ghose, & A. Subedi, "Deep learning approach on network intrusion detection system using NSL-KDD dataset," International Journal of Computer Network and Information Security, vol. 11, no. 3, pp. 8-14, 2019.

[75] M. Nkongolo, J.P. van Deventer, & S.M. Kasongo, "UGRansome1819: A Novel Dataset for Anomaly Detection and Zero-Day Threats," Information, vol. 12, no. 10, pp. 405, 2021.

[76] S. Shunmuganathan, R.D. Saravanan, & Y. Palanichamy, "Securing VPN from insider and outsider bandwidth flooding attack," Microprocessors and Microsystems, vol. 79, pp. 103279, 2020.

[77] S. Saraswathi, V. Krishnamurthy, D.V.V. Prasad, R.K. Tarun, S. Abhinav, & D. Rushitaa, "Machine learning based Ad-click prediction system," Int J Eng Adv Technol, vol. 8, no. 6, pp. 3646-3648, 2019.

[78] S. Shunmuganathan, "A Reliable Lightweight Two Factor Mutual Authenticated Session Key Agreement Protocol for Multi-Server Environment," Wireless Personal Communications, vol. 121, no. 4, pp. 2789-2822, 2021.

[79] R.D. Saravanan, S. Loganathan, S. Shunmuganathan, & Y. Palanichamy, "Suspicious score based mechanism to protect web servers against application layer distributed denial of service attacks," Int J Intell Eng Syst, vol. 10, no. 4, pp. 147-156, 2017.

[80] S.R. Devi, S. Saraswathi, & P. Yogesh, "A Cooperative Multilayer End-Point Approach to Mitigate DDoS Attack," WSEAS Transactions on Information Science and Applications, vol. 11, no. 1, pp. 1-11, 2014.

[81] R. Saravanan, S. Shanmuganathan, & Y. Palanichamy, "Behavior-based detection of application layer distributed denial of service attacks during flash events," Turkish Journal of Electrical Engineering & Computer Sciences, vol. 24, no. 2, pp. 510-523, 2016.

[82] R. Singh, H. Kumar, R.K. Singla, & R.R. Ketti, "Internet attacks and intrusion detection system: A review of the literature," Online Information Review, 2017.