International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 8 Issue 4, Jul-Aug 2024 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Detail Analysis of Attacks and Methods of Intrusion Detection System

Keshav Sinha[1], Partha Paul[2]

[1]Department of Computer Science & Engineering,
University of Petroleum & Energy Studies, Dehradun, Uttarakhand, India
[2]Department of Computer Science & Engineering, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India

ABSTRACT

Computer networks link many activities, events, and applications. The network's performance must be improved and have more capacity to handle increased users. The network's computer system should guarantee security, confidentiality, and integrity. An intrusion jeopardizes the operation and security of a wired or wireless network system. If the invasions are not detected at the appropriate level, the loss to the system might be immeasurable. Intrusions occur when malicious actors harm information resources. The hackers tamper the normal operations or attempt to infiltrate the system via the gateway. The study analyzes the attack and normal traffic packets from the KDD Cup99 dataset. The KDD Cup99 data includes benchmark traffic and intrusion detection features. However, most intrusion detection systems today have significant false alarm rates and miss many attacks because they cannot distinguish between unlawful and unlawful behaviors.

KEYWORDS: *Invasion, Intrusion Detection, KDDCup99, Misuse detection, Anomaly detection*

How to cite this paper: Keshav Sinha | Partha Paul "Detail Analysis of Attacks and Methods of Intrusion Detection System" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-8 | Issue-4, August 2024, pp.1032-1041, URL: www.ijtsrd.com/papers/ijtsrd68276.pdf

IJTSRD68276

I. INTRODUCTION

Daily life depends on the availability and processing of information quickly. If demand increased in this scenario, it would be necessary to store proportionately more data and resources across numerous computers with the necessary correlation, and data interference, unauthorized access, and system and network growth would worsen. The virtual access path would grant access to unauthorized network users. On the other side, hackers can access confidential data by taking advantage of flaws in networks or systems. The constraints on access and security measures are insufficient against internal and compromised threats. Recognizing breaches and intrusions is the only proven approach to keeping systems and networks safe. Along with identifying real attackers, intrusion detection systems should also keep track of attempted intrusions.

A trustworthy system should secure its data and resources from unauthorized access, tampering, and denial of service attacks. The function of any computer network system should have some expected level of trust and confidence. The security policy must be formulated for every system based on future performance. Computer security is typically based on realizing the following factors in a computer system.

- **Confidentiality:** information is to be accessed only by authorized persons.
- **Integrity:** information must remain unaltered by mischievous or malicious attempts.
- **Availability:** the computer must function without degradation of access and impart resources to legitimate users when required.

In general, an intrusion is any action attempting to compromise a resource's confidentiality, integrity, and availability. Anderson (1980) defined intrusion as the potential opportunity of an intentionally unauthorized attempt to access information, manipulate information, or make a system untrustworthy.

Intrusion Detection System (IDS) was commercially introduced in the year 1990 [1]. It behaves like a burglar alarm that detects invasion and triggers alarms like audible, visual, or messages like e-mail. The IDS is used to prevent problem behaviors that attack or abuse the system, detect, and deal with attacks. The mechanism should have low false alarms while ensuring invasion detection. Various approaches are present, but they are relatively ineffective in the classification and alarm rate dimensions. Machine learning-based anomaly detection approaches have been effectively used in the network intrusion detection scenario because of their intrinsic capabilities of discovering new attacks [2]. Most existing classification methods are based on neural networks, fuzzy logic, genetic algorithm, and support vector machines.

The motivation for this work is to study and analyze various attacks and explore benchmark datasets for designing the enhanced methodology. The paper's objective is to study the existing available methods to explore the possibilities of improved performance. The rest of the paper is organized as follows: Section 1 presents the Introduction to IDS, Section 2 presents the surveys of significant work carried out in the domain of IDS, Section 3 describes the analysis of data applicable to IDS, and finally, the conclusion is presented for the entire work.

II. RESEARCH BACKGROUND

Detection of intrusions protects a computer network from unauthorized users as well-as insiders attack. The intrusion detector task is to construct a predictive model or classification method capable of distinguishing 'bad' connections, called intrusions or attacks, and 'good' or average connections. IDSs are broadly classified into three categories based on deployment.

- **Network-based IDS (NIDS):** It is a passive device that resides in an organization's computer or network and observes the network traffic to indicate attacks. It recognizes any attack and notifies such malicious codes to system administrators immediately. It can be installed in the boundary of the router to observe the traffic going into and out of the network [3]. The minimum number of monitoring units for an extensive network can be deployed without disturbing the regular operations of networks. It is also not vulnerable to direct attack, but it can become exhausted by network traffic, unable to detect encrypted packets and fail to distinguish some attacks.
- **Host-based IDS (HIDS):** It resides in the computer or server, called the host, and examines only the host activities. It is employed to monitor the system and stored configuration files and detect the intruders' creation, modification, and deletion of system files. It can also detect local events and attacks that the NIDS has not detected. The configuration of HIDS resides only on an individual host and requires more management effort to install and configure in multiple hosts. Also, HIDS are more vulnerable to direct attacks and susceptible to some Denial of Service (DoS) attacks [4].
- **Application-based IDS (AppIDS):** It is the enhancement of the HIDS, which examines an application for abnormal events by looking into the files created in the application and anomaly occasions such as exceeding the users' authorization, and void file execution. It also observes the interaction between the application and the user and the encrypted traffic. It is more susceptible to attack and does not possess the skill to detect software tampering [5].

The accuracy of any IDS is measured based on the false alarm rate (both positive and negative). Based on the detection method, IDSs are classified into:

- **Misuse Detection:** In misuse detection or signature- based intrusion detection system, the signatures or patterns of the known attacks are placed in the database. They are matched with the signatures of traffic entering the network. In case of any attack, the signature can be used to detect it accurately. Unfortunately, newly formed attacks with modified signatures can go undetected within the system and are classified as false negatives [6]. In general, many false negatives are more associated with signature-based IDS. It is also referred to as knowledge-based IDS.
- **Anomaly Detection:** The anomaly detection or statistical anomaly-based IDS gathers statistical summaries by watching the traffic, which is known to be expected, and a performance baseline is developed. The network activities are periodically monitored and compared with the baseline of intrusions. The statistical and behavioral patterns that detect attacks allow a low false negative rate. The behavioral patterns of users or programs are used to develop a pattern of normal and abnormal activities, which are used to detect the occurrence of an attack. Consequently, any variation from typical behavior by a user or program would be detected, thereby generating an alarm. Regrettably, most alarms are benign and false positives are derived as a result. It is also referred to as behavior-based IDS [7].

The fundamental principle of anomaly intrusion detection is that any intrusive activity is a subset of bizarre action. The intrusion may be recognized based on anomalous actions. For example, suppose an authorized employee of an organization opens the system after office hours using their official account. In that case, it is also considered abnormal, and consequently, it may be an intrusion. Likewise, users in an organization constantly login out of working hours through the official server is also treated as an anomaly. The intrusive activity can be carried out as a sum of individual activities, and no one is separately anomalous. Flagging every part of anomalous activities precisely results in false positives or false negatives. However, intrusive activity does not coincide all the time with anomalous activity. There are four possibilities (Sangeetha et al., 2022):

- **Intrusive but not Anomalous**: It is also called false negatives or Type I errors, in which the activity is intrusive and fails to detect because it is not anomalous. These are false negatives because the IDS falsely reports the absence of intrusions.
- **Not Intrusive but Anomalous:** It is also called false positives or Type II errors, in which the activity is not intrusive and treated as intrusive because it is anomalous. These are called false positives because the IDS falsely reports intrusions.
- **Not Intrusive and not Anomalous:** It is also called true negatives, in which the activity is not intrusive and is not informed as intrusive.
- **Intrusive and Anomalous:** It is also called true positives, in which the activity is intrusive and reported as intrusive because it is also anomalous.

In an anomaly detection system, the activities of various subjects are observed, and profiles are generated based on behaviors called master profiles. If any behavior changes happen in the upcoming period, the new profile measures will be updated periodically. The current activities are stored in temporary profiles and periodically transferred to a master profile. In statistical intrusion detection systems, acquiring user activities would be trained regularly using behavioral moment, which is used to distinguish the patterns as normal or abnormal. The advantage of anomaly intrusion detection is that the data point of a specific feature that lies away from a multiple of the standard deviation (statistics) on both sides of the mean may be measured as anomalous. The disadvantages are anomaly intrusion detections are not sensitive to the order of incidence of events. They will probably miss intrusions that are indicated by sequential interrelationships among events. Moreover, fixing the threshold value of deviation is challenging—the shallow threshold setting results in false positives, and high-value results in false negatives.

The false positives are the provocation of intrusion detection systems. Anomaly detection systems are mainly prone to false positives. Generally, no significant rate of false positives in signature-based systems is reported if rules are correctly installed. Likewise, false negatives are also a problem for IDS. Typical data may generate false negatives in misuse-based systems due to the resemblance of existing attacks. The techniques for detecting intruders have evolved to face new attacks. It simplifies that the standard and attack packets are indicated by '0' and '1', respectively. **Table 1** presents various works in terms of security systems and feature selection. The Hybrid Association Classification (AC) approach, a hybrid classification methodology, was introduced by Hadi et al., (2018). Several rules are developed to reflect each attribute, and the number of categorization rules is maintained to a minimum. Two Extreme Layer Machines (TELM) were suggested by Qu et al., (2016) to tackle challenging classification and regression problems with little storage. When a neural network has a lot of hidden layers, TELM significantly improves performance. Nabipour et al. (2020), proposed a classification approach for high-dimensional situations. The genetic algorithm supports the fuzzy rule-based methodology used to create the classification model. The guidelines for choosing the best features were predicted using the Mixed Integer Programming Model.

Table 1. Chronological Literature Review

Research	Technique Used	Methodology	Advantages/Disadvantages
Nadiamm ai *et al.* (2014)	• Intrusion Detection System (IDS) with Data Mining. • Efficient Data Adapted Decision Tree (EDADT).	• Identify the relevant data. • Classify the Distributed Denial of Service (DDoS) attack using labeled data.	• Efficient • High Detection Rate • High Accuracy
Shakshuki et al., (2012)	• IDS for MANET	• Enhanced Adaptive Acknowledgment (EAACK) • Classify the malicious behavior	• High Detection Rate • Low False Alarm Rate
Bhatia et al., (2017)	• IDS with Artificial Neural Network (ANN)	• Classify the training data • Compare the oversampling of the U2R and R2L • Categories the attacks	• Better Detection Rate • High Accuracy

Yahalom et al., (2019)	• Intrusion Detection System for Hierarchical Data	• MIL-STD protocol is used for the training of data • Reduce the false alarm rate	• High Detection Rate • Efficiency • High Accuracy
Liu & Lang (2019)	• Intrusion Detection by fusion of different feature selection algorithms.	• Feature Selection is performed using Linear correlation coefficient and Cuttlefish algorithms	• High Detection Rate • Low False Alarm Rate • High Accuracy

One of the best approaches to solving the multi-class problem in machine learning is to use a classification system based on fuzzy rules. The Cluster Center and Nearest Neighbor feature selection method was put out by Lin et al. (2015). It computes the distance between each data sample and its own cluster's center by calculating the distance and then using the same function on the data and the cluster's closest neighbor.

Then, using the k-NN classifier, which has a high processing efficiency and detection rate, each piece of data may be utilized in the intrusion detection process. Composition of Feature Relevancy is a novel feature selection method proposed by Longde et al., (2018). The eight real-world datasets and two different classifiers are used to enhance feature selection. Liu et al., (2017) proposed a technique for selecting attributes based on aptitude. After identifying the closest traits, the quality is determined. These methods result in superior feature selection outcomes. Basu (2019) invented a brand-new data structure called a Grid Count Tree (GCD) to find outliers. It may be used to compute numerical value separation and category separation quickly and to separate meaningful signals from false data. Both real- world and artificial genetically connected applications are used to evaluate this GCD. Cai (2013) introduced the Iterative Self-organizing Map with Robust Distance (ISOMRD) for outlier detection based on this situation. When points with similar traits assemble, clusters are created. Many databases are processed via iterative processing. It is helpful to locate solutions for dynamic analysis and geographical data mining applications. Bai et al., (2016) proposed an outlier detection technique based on the local outlier factor for large data sets. Outliers are identified using the Grid-Based Partition Algorithm and the Distributed LOF. The data collection is divided into a small number of grid sets, and data nodes are assigned. Tuples are categorized using classification as cross- grid tuples or gird local tuples. Dispersed LOF is utilized effectively in distributed situations to reduce outliers. Di Mauro et al. (2021), suggested a feature selection technique for two categories of data sets. These data sets to aid in the detection of false negatives and improve forecast accuracy. Idris (2014) suggested a negative feature selection technique for detecting e-mail spam. NSA- PSO defines a local outlier factor to estimate the threshold value. The proposed method outperforms non-FSA techniques. Under the title Distribution Estimation based Negative Selection Algorithm, Fouladvand et al., (2017) introduced a novel attribute selection approach for normal and self-space using detectors (DENSA). Random detectors performed well on a range of real-world data sets in this experiment.

Various applications exist for the IDS to detect different types of attacks and security violations. It also prevents the applications such as Business transaction systems, Document maintenance systems, Banking, Insurance Systems, and E-Governance from the adversary. The applications of IDS are not specified because all sensitive services are available on the Internet and Intranet. The service providers need to safeguard valuable information consistently. The technologies are growing exponentially, and protecting resources is becoming more complex. The system framework and the critical elements of the research model are covered in the following section.

III. THEORETICAL FRAMEWORK

Real-world data must be generated for intrusion detection to evaluate all potential risks. The stages involved in data analysis methodically identify patterns in the gathered data and link them to the problem that has been recognized. Data modeling will determine how it may be categorized and connected. The accuracy and reliability of the data collected for the evaluation are aspects of data quality. **Figure 1** presents the theoretical Framework for KDD Cup Dataset Analysis. The investigation would be feasible if the data quality and attributes for the position were excellent.

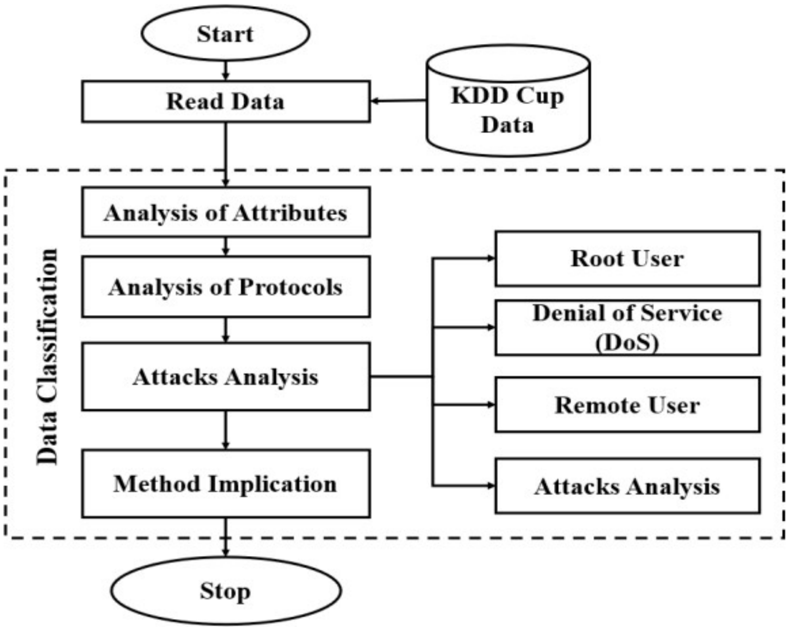
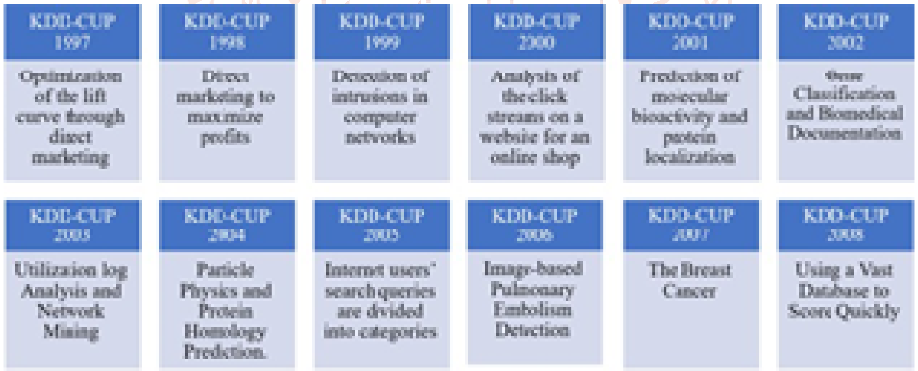


Figure 1. Theoretical Framework for KDD Cup Dataset Analysis

The research requires several ground truth databases in its region. In this paper, KDD Cup99 is used for intrusion detection systems to classify network traffic. The dataset consists of professional-level interest groups on knowledge discovery and data mining (http://www.sigkdd.org/kddcup) (Nguyen et al., 2016). The Lincoln Laboratory at Massachusetts Institute of Technology produced standard network traffic data under the auspices of DARPA and the Air Force Research Laboratory to evaluate computer network intrusion. The research activity mainly focuses on the 1998 and 1999 datasets. **Figure 2** presents the KddCup99 dataset description.



KDD-CUP 1997	KDD-CUP 1998	KDD-CUP 1999	KDD-CUP 2000	KDD-CUP 2001	KDD-CUP 2002
Optimization of the lift curve through direct marketing	Direct marketing to maximize profits	Detection of intrusions in computer networks	Analysis of the click streams on a website for an online shop	Prediction of molecular bioactivity and protein localization	Gene Classification and Biomedical Documentation

KDD-CUP 2003	KDD-CUP 2004	KDD-CUP 2005	KDD-CUP 2006	KDD-CUP 2007	KDD-CUP 2008
Utilization log Analysis and Network Mining	Particle Physics and Protein Homology Prediction.	Internet users' search queries are divided into categories	Image-based Pulmonary Embolism Detection	The Breast Cancer	Using a Vast Database to Score Quickly

Figure 2: KddCup99 Dataset Description

A standardized set of auditable data containing a variety of simulated intrusions data present on the military network environment. It emulated on US Air Force LAN, mainly focused on real environment attacks. The raw TCP/IP dump data has been captured from the network. A connection between the source IP and destination IP address is presented in TCP sequence packets. It starts and stops at certain times that allow data to transfer per specific protocol. Furthermore, each connection contains a label that indicates normal or an assault with a specific attack type.

A. Attributes in KddCup99

The features are grouped into three categories (i) basic features of individual connections, (ii) content features within a connection, and (iii) traffic features which are computed using a two-second time. The KDD Cup99 uses a series of packets with a total of 41 characteristics that are broadcast over two seconds. A packet's fundamental features are represented by features (0-9), content features are represented by features (10-22), traffic features are represented by (23-31), and host-based features from (32–41). Some of the terminologies associated with the data set are (i) Connections that were established with the same host as the one being utilized for the current connection within the previous two seconds are referred to as having the 'same host,' and (ii) the term 'same service' refers to connections that provided the same service as the one being used now within the last two seconds. The characteristics based on 'same host' and 'same service' are collectively referred to as the time-based traffic aspects of the connection records. **Table 2** present the various attributes of KddCup99 datasets.

Table 2. KddCup99 Attribute Description

Feature Name	Variable	Type	Label	Description
duration	C	1	v1	Connections in seconds
protocol_type	D	1	v2	Types of protocol (TCP, UDP, etc.)
service	D	1	v3	Network service (HTTP, telnet, etc.)
flag	D	1	v4	Normal or Error connection status.
src_bytes	C	1	v5	Source to Destination data bytes info
dst_bytes	C	1	v6	Destination to Source data bytes info
land	D	1	v7	1-Connection from/to host/port. 0-otherwise
wrong_frag ment	C	1	v8	Number of 'wrong' fragments
urgent	C	1	v9	Number of urgent packets
hot	C	2	v10	Count the System Access
num_failed_logins	C	2	v11	Number of failed login attempts
logged_in	C	2	v12	1-Successfully logged; 0- otherwise
num_comp romised	C	2	v13	Compromised conditions
root_shell	C	2	v14	1 - root shell is obtained; 0 - otherwise
su_attempted	C	2	v15	1-SU root 0 - Otherwise
num_root	C	2	v16	'Root' accesses
num_file_c reations	C	2	v17	File creation operations
num_shells	C	2	v18	Number of shell prompts
num_acces s_files	C	2	v19	Writes, delete and create operations.
num_outbo und_cm ds	C	2	v20	Outbound commands in FTP
is_hot_login	D	2	v21	1-Login 'hot' list (root, adm, etc.); 0-otherwise
is_guest_login	D	2	v22	1-Login (guest, anonymous, etc.); 0-otherwise
count	C	3	v23	Same Host Connections
srv_count	C	3	v24	Connections to same Service
serror_rate	C	3	v25	'SYN' errors to the same host
srv serror rate	C	3	v26	'SYN' errors to the same service
rerror_rate	C	3	v27	'REJ' errors to the same host
srv_rerror_rate	C	3	v28	'REJ' errors to the same service
same_srv_rate	C	3	v29	Same Service and the same host
diff_srv_rate	C	3	v30	Different Services and the same host
srv_diff_ho st_rate	C	3	v31	Same Service and different hosts
dst_host_count	C	3	v32	Same Host to the Destination Host
dst_host_srv_count	C	3	v33	Same Service to Destination Host as Current Connection
dst_host_sa me_srv_rate	C	3	v34	Same Service to the Destination Host
dst_host_di ff_srv_rate	C	3	v35	Different Services to the Destination Host
dst_host_same_src_port_rate	C	3	v36	Port Services to the Destination Host
dst_host_srv_diff_host_rate	C	3	v37	Different Hosts from the same service to the destination host
dst_host_se rror_rate	C	3	v38	'SYN' (errors same host to destination)
dst_host_srv_serror_rate	C	3	v39	'SYN' errors from the same service to the destination host
dst_host_rerror_rate	C	3	v40	'REJ' errors (same host to destination)
dst_host_sr v_rerror_rate	C	3	v41	'REJ' errors (same service to destination)

**** C- Continuous, D- Discrete**1-Intrinsic, 2-Content, 3-Traffic***

The protocol_type, service, flag, land, logged_in, is_hot_login, and is_guest_login is labeled as discrete or categorical features, and the other 34 features are labeled as continuous features. **Table 3** present the description of various flag values of KddCup99, and the categorical features protocol_type, service, and flag have different values listed in **Table 4**.

Table 3. Description of flag values

Flag	Label	Description
RSTOS0	1	The originator sent an SYN followed by an RST but never see an SYN-ACK from the responder
RSTR	2	Established, responder aborted
RSTO	3	Connection established; originator aborted (sent an RST)
OTH	4	No SYN seen, just midstream traffic (a "partial connection" that was not later closed)
REJ	5	Connection attempt rejected
S0	6	A connection attempt was seen, but no reply
S1	7	Connection established, not terminated
S2	8	Connection established and the close attempt by originator seen (but no reply from responder)
S3	9	Connection established and the close attempt by responder seen (but no reply from originator)
SF	10	Normal establishment and termination
SH	11	The originator sent an SYN followed by a FIN (finish 'flag') but never saw an SYN-ACK from the responder (hence the connection was "half" open)

Table 4. Various services and flags in the KddCup99 dataset

Label	Service	Label	Service	Label	Service
1	netbios_dgm	25	Z39_50	49	time
2	netbios_ssn	26	gopher	50	echo
3	netbios_ns	27	domain	51	ldap
4	remote_job	28	finger	52	link
5	http_8001	29	klogin	53	HTTP
6	hostnames	30	kshell	54	SMTP
7	uucp_path	31	supdup	55	UUCP
8	http_2784	32	systat	56	auth
9	iso_tsap	33	telnet	57	nnsp
10	csnet_ns	34	shell	58	nntp
11	domain_u	35	imap4	59	name
12	ftp_data	36	eco_i	60	exec
13	http_443	37	ecr_i	61	AOL
14	daytime	38	red_i	62	IRC
15	harvest	39	pop_2	63	X11
16	discard	40	pop_3	64	BGP
17	netstat	41	login	65	CTF
18	courier	42	tim_i	66	MTP
19	pm_dump	43	urh_i	67	rje
20	printer	44	urp_i	68	ssh
21	private	45	ntp_u	69	efs
22	sql_net	46	vmnet	70	ftp
23	tftp_u	47	other		
24	sunrpc	48	whois		

B. Classification of Attacks

There are varieties of attacks which are entering into the network over a period, and the attacks are classified into the following four main classes:

- ***Denial of Service:*** It is a class of attacks where an attacker makes some computing or memory resource too busy or too full to handle legitimate requests, denying legitimate users access to a machine. The three different ways to launch a DoS attack are (i) by abusing the computer's legitimate features, (ii) by targeting the implementation bugs, and (iii) by exploiting the misconfiguration of the systems. The DoS attacks are classified based on the services an attacker renders unavailable to legitimate users.

- ***User to Root:*** The attacker starts with access to a normal user account on the system and gains root access. Common programming mistakes and environment assumptions allow attackers to exploit root access's vulnerability.
- ***Remote to User:*** The attacker sends packets to a machine over a network that exploits the machine's vulnerability to gain local access as a user illegally. There are different types of R2L attacks, and the most common attack in this class is made using social engineering.
- ***Probing:*** It is a class of attacks where an attacker scans a network to gather information to find known vulnerabilities. An attacker with a map of machines and services available on a network can manipulate the information to look for exploits. Different probes exist; some abuse the computer's legitimate features, and some use social engineering techniques.

Table 5 present the various class of attacks that is most common for the analysis of the KddCup99 dataset.

Table 5. Various attacks on KddCup99 Dataset

Attack	Type	Mechanism	Attack Effect
back	DoS	Abuse/Bug	Slows down server response
land	DoS	Bug	Slows down server response
Neptune	DoS	Abuse	Slows down server response
smurf	DoS	Abuse	Slows down the network
pod	DoS	Abuse	Slows down server response
teardrop	DoS	Bug	Reboots the machine
load- module	U2R	Poor environment sanitation	Gains root shell
buffer_over flow	U2R	Abuse	Gains root shell
rootkit	U2R	Abuse	Gains root shell
Perl	U2R	Poor environment sanitation	Gains root shell
phf	R2L	Bug	Executes commands as root
guess_pass wd	R2L	Login misconfiguration	Gains user access
warezmaste r	R2L	Abuse	Gains user access
IMAP	R2L	Bug	Gains root access
multihop	R2L	Abuse	Gains root access
ftp_write	R2L	Misconfigura tion	Gains user access
spy	R2L	Abuse	Gains user access
warezclient	R2L	Abuse	Gains user access
satan	Probe	Abuse of feature	Looks for known vulnerabilities
Nmap	Probe	Abuse of feature	Identifies active ports on a machine
portsweep	Probe	Abuse of feature	Identifies active ports on a machine
ipsweep	Probe	Abuse of feature	Identifies active machines

The data set in KDD Cup99 have normal, 22 attack-type data with 41 features, and **Table 6** shows a few data set. All generated traffic patterns end with a label either as 'normal' or any 'attack' for upcoming analysis.

Table 6. Sample Data Packets

Feature Name	Packet-1 (Normal)	Packet-2 (Neptune)

duration	0	0
protocol_type	TCP	TCP
service	HTTP	private
Flag	SF	REJ
src_bytes	327	0
dst_bytes	467	0
Land	0	0
wrong_fragment	0	0
urgent	0	0
Hot	0	0
num_failed_logins	0	0
logged_in	1	0
num_compromised	0	0
root_shell	0	0
su_attempted	0	0
num_root	0	0
num_file_creations	0	0
num_shells	0	0
num_access_files	0	0
num_outbound_cmds	0	0
is_hot_login	0	0
is_guest_login	0	0
count	33	136
srv_count	47	1
serror_rate	0	0
srv_serror_rate	0	0
rerror_rate	0	1
srv_rerror_rate	0	1
same_srv_rate	1	0.01
diff_srv_rate	0	0.06
srv_diff_host_rate	0.04	0
dst_host_count	151	255
dst_host_srv_count	255	1
dst_host_same_srv_rate	1	0
dst_host_diff_srv _rate	0	0.06
dst_host_same_src_port_rate	0.01	0
dst_host_srv_diff_host_rate	0.03	0
dst_host_serror_rate	0	0
dst_host_srv_serror_rate	0	0
dst_host_rerror_rate	0	1
dst_host_srv_rerror_rate	0	1

This section outlines the structure of the dataset used by the Intrusion detection system. The various kinds of features, such as discrete and continuous, are studied with a focus on their role in the attack. The attacks are classified with a brief introduction to each.

IV. CONCLUSION

Any network administrator's primary priority should be intrusion detection. We conducted a thorough yet simple study to examine different methods for developing Network Intrusion Detection models. Several research articles published in various journals served as the foundation for the construction of this study. Several tables provided in this publication analyze the Kddcup99 dataset's characteristics. The many strategies employed by the network intrusion detection system are described, along with each one's benefits and drawbacks. It also observed the presence of many assault packets, both normal and attack. The investigation in this work is broadened based on several machine learning methods for identifying the system assault. While the machine is given the ability to learn, the behavior of the data has been studied for further research.

REFERENCES

[1] Bass, T. (2000). Intrusion detection systems and multisensor data fusion. Communications of the ACM, 43(4), 99-105.

[2] Rimmer, V., Nadeem, A., Verwer, S., Preuveneers, D., & Joosen, W. (2022). Open-World Network Intrusion Detection. In Security and Artificial Intelligence (pp. 254-283). Springer, Cham.

[3] Horchulhack, P., Viegas, E. K., & Santin, A. O. (2022). Toward feasible machine learning model updates in network-based intrusion detection. Computer Networks, 202, 108618.

[4] Ahmet, E. F. E., & ABACI, İ. N. (2022). Comparison of the Host Based Intrusion Detection Systems and Network Based Intrusion Detection Systems. Celal Bayar University Journal of Science, 18(1), 23-32.

[5] Agarwal, N., & Hussain, S. Z. (2018). A closer look at intrusion detection system for web applications. Security and Communication Networks, 2018.

[6] Fernando, P., Dadallage, K., Gamage, T., Seneviratne, C., Madanayake, A., & Liyanage, M. (2022). Proof of Sense: A Novel Consensus Mechanism for Spectrum Misuse Detection. IEEE Transactions on Industrial Informatics, 18(12), 9206-9216.

[7] Sinha, K., & Verma, M. (2021). The Detection of SQL Injection on Blockchain-Based Database. In Revolutionary Applications of Blockchain-Enabled Privacy and Access Control (pp. 234-262). IGI Global.

[8] Sangeetha, S. K., Mani, P., Maheshwari, V., Jayagopal, P., Sandeep Kumar, M., & Allayear, S. M. (2022). Design and Analysis of Multilayered Neural Network-Based Intrusion Detection System in the Internet of Things Network. Computational Intelligence & Neuroscience, 2022.

[9] Hadi, W. E., Al-Radaideh, Q. A., & Alhawari, S. (2018). Integrating associative rule-based classification with Naïve Bayes for text classification. Applied Soft Computing, 69, 344-356.

[10] Qu, B. Y., Lang, B. F., Liang, J. J., Qin, A. K., & Crisalle, O. D. (2016). Two-hidden-layer extreme learning machine for regression and classification. Neurocomputing, 175, 826-834.

[11] Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., & Mosavi, A. (2020). Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis. IEEE Access, 8, 150199-150212.

[12] Nadiammai, G. V., & Hemalatha, M. J. E. I. J. (2014). Effective approach toward Intrusion Detection System using data mining techniques. Egyptian Informatics Journal, 15(1), 37-50.

[13] Shakshuki, E. M., Kang, N., & Sheltami, T. R. (2012). EAACK—a secure intrusion-detection system for MANETs. IEEE Transactions on industrial electronics, 60(3), 1089-1098.

[14] Bhatia, M. K., Ripudaman, S., Akashdeep, S., & Bhardwaj, B. L. (2017). Knowledge, Attitude and Practice of self-medication among undergraduate medical students of Punjab. J Med Res, 3(3), 151-4.

[15] Yahalom, R., Steren, A., Nameri, Y., Roytman, M., Porgador, A., & Elovici, Y. (2019). Improving the effectiveness of intrusion detection systems for hierarchical data. Knowledge-Based Systems, 168, 59-69.

[16] Liu, H., & Lang, B. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. applied sciences, 9(20), 4396.

[17] Lin, W. C., Ke, S. W., & Tsai, C. F. (2015). CANN: An intrusion detection system based on combining cluster centers and nearest neighbors. Knowledge-based systems, 78, 13-21.

[18] Longde, S. U. N., Xiaolin, W. U., Wanfu, Z. H. O. U., Xuejun, L. I., & Peihui, H. (2018). Technologies of enhancing oil recovery by chemical flooding in Daqing Oilfield, NE China. Petroleum Exploration and Development, 45(4), 673-684.

[19] Liu, J., Lin, Y., Lin, M., Wu, S., & Zhang, J. (2017). Feature selection based on quality of information. Neurocomputing, 225, 11-22.

[20] Basu, P. (2019). Toward Reliable, Secure, and Energy-Efficient Multi- Core System Design (Doctoral dissertation, Utah State University).

[21] Cai, Q. (2013). Self-organizing learning model for data mining applications. Stevens Institute of Technology.

[22] Bai, M., Wang, X., Xin, J., & Wang, G. (2016). An efficient algorithm for distributed density-

based outlier detection on big data. Neurocomputing, 181, 19-28.

[23] Di Mauro, M., Galatro, G., Fortino, G., & Liotta, A. (2021). Supervised feature selection techniques in network intrusion detection: A critical review. Engineering Applications of Artificial Intelligence, 101, 104216.

[24] Idris, I., & Selamat, A. (2014). Improved email spam detection model with negative selection algorithm and particle swarm optimization. Applied Soft Computing, 22, 11-27.

[25] Fouladvand, S., Osareh, A., Shadgar, B., Pavone, M., & Sharafi, S. (2017). DENSA: An effective negative selection algorithm with flexible boundaries for self-space and dynamic number of detectors. Engineering Applications of Artificial Intelligence, 62, 359-372.

[26] Nguyen, T. T., Nguyen, T. T. T., Pham, X. C., & Liew, A. W. C. (2016). A novel combining classifier method based on variational inference. Pattern Recognition, 49, 198-212.