Archives available at journals.mriindia.com

International Journal on Advanced Electrical and Computer Engineering

ISSN: 2349-9338

Volume 15 Issue 01s, 2026

Intelligent Intrusion Detection Systems Using Machine Learning

[1]Mrs.Rajashri Santosh Shekokar , [2]Dr.Krishnakant.P.Adhiya

Peer Review Information

Submission: 05 Dec 2025

Revision: 25 Dec 2025

Acceptance: 10 Jan 2026

Keywords

IDS, AI, anomaly & misuse, DOS, R2L, NID, Firewall.

Abstract

Intrusion Detection System (IDS) watches network traffic for fraudulent activity and gives immediate alerts when it is observed. By notifying security administrators of known or possible threats or by sending alerts to a centralized security tool, an intrusion detection system (IDS) can assist speed up and automate network threat detection. This paper presents a comprehensive overview of Intrusion Detection Systems (IDS) and the machine learning techniques commonly employed to enhance their detection capabilities. Various IDS approaches are examined, with a focus on both signature-based and anomaly-based models. The study highlights key machine learning methods—such as Support Vector Machines (SVM), Convolutional Neural Networks (CNN), Random Forests, and clustering algorithms—and evaluates their suitability for identifying malicious activities in network environments. The growing integration of AI in cybersecurity is also discussed, emphasizing its role in improving automated threat analysis and adaptive intrusion detection Additionally, the KDD-99 dataset is utilized to outline experimental procedures and demonstrate how these algorithms can be applied in practical IDS implementations. The findings emphasize the importance of selecting appropriate learning techniques to improve accuracy, reduce false positives, and strengthen overall network security.

Introduction

A network security technology called an intrusion detection system (IDS) keeps an eye on devices and network traffic for known hostile activities, suspicious activity, or infractions of security policies. IDSs can be specialized hardware devices connected to the network or software programs deployed on endpoints. A signature-based IDS maintains a database of attack signatures against which it checks network packets. Anomaly-based detection methods use machine learning to establish, and continually update, a baseline model of normal network behavior. IDSs are classified according to the type of activity they monitor and their placement within a system. Network intrusion detection systems (NIDSs) monitor inbound and outbound traffic to devices across the network. Host intrusion detection systems (HIDSs) are installed on a specific endpoint, like a laptop, router, or server. A protocol-based intrusion detection system is commonly implemented on a web server. It keeps an eye on and examines the protocol that connects a user or device to the server. An APIDS is a mechanism or agent that usually sits inside the server party. It tracks and interprets correspondence on application-specific protocols. A hybrid intrusion detection system incorporates two or more intrusion detection technologies. utilizing network data in conjunction with system or host agent data to provide a thorough understanding of the system. The potential thread to the global information infrastructure is growing as computer networks and their associated applications become more and more common in the information society era

[1]. Yahoo became the first target of a denial-of-service assault in 2000, the same year that DOS made its first attack public. At the present time, digital services and social websites are target of DOS attacks [2]. Denial-of-service attacks under a number of guises have been around for decades. Distributed DoS attacks are much newer, first being seen in late June and early July of 1999 [3].

Working Of IDS :
An intrusion detection system (IDS) monitors network traffic for any unexpected activity. It checks the data going over the network for patterns and indicators of anomalous activity. The IDS compares network traffic to a set of predefined rules and patterns to detect any activity that may indicate an attack or intrusion. If the IDS detects something that matches one of these rules or patterns, it alerts the system administrator.
The system administrator can then evaluate the alert and take necessary steps to prevent future harm or infiltration.

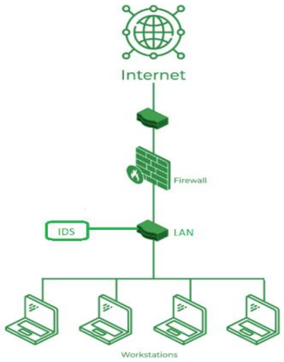

Fig 1 : Intrusion Detection System (Source : Internet)

Types of IDS and how they work Network-based IDS (NIDS).

1. Host-based IDS (HIDS)
Host-based IDS Operates on specific devices or hosts, monitoring actions within the operating system and apps.
It Analyzes logs, file system modifications, and system calls to detect suspicious behavior or unauthorized access.Basically useful in detecting insider attacks and malware activities.Following Diagram Shows Working Of HIDS.

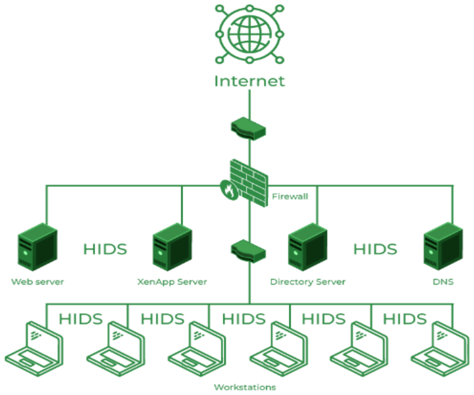

Fig 2 : Showing Working of HIDS. (Source : Internet)

Advantages of HIDS

1. HIDS Predict What An Application Does.
2. HIDS Helps to Detect Attacks Excluded From the Network.

Disadvantage of HIDS

1. HIDS Can be excluded from the network.
2. HIDS can be Installed on all hosts, this passive system only alerts to attacks without taking action.

Techniques used: Autoencoders , Clustering (e.g., k-means),One-class SVM ,Isolation forests

2. Signature-based IDS
Signature-Based IDS Identifies hazards by referring to a database of known attack patterns or signatures. To find matches, this tool compares incoming traffic or system activity to certain signatures. Effective against known threats, but may struggle to identify novel or unexpected attacks.

Advantages of Signature-based IDS

1. Signature Based IDS shows High accuracy against known threats: Detects and prevents known attacks by comparing communications to a database of predefined harmful patterns.
2. Low false positive rate: Because signatures are static and based on recognized dangers, they are less likely to generate false alarms than other approaches.

Disadvantages of Signature-based IDS

1. Signature Based IDS is Ineffective against new threats: cannot identify "zero-day" assaults or variants of known malware for which no signature is yet available.
2. Signature Based IDS Requires constant updates: To remain effective, the signature database must be regularly and

continuously updated, which is a resource-intensive maintenance task.

Techniques : Deep Neural Networks (DNN), Convolutional Neural Networks (CNN),Decision Trees / Random Forests

Anomaly-based IDS

Anomaly-Based IDS Establishes a baseline of "normal" behavior and flags deviations as potential threats. learns what constitutes typical behavior and notifies users when actions substantially deviate from the predetermined baseline. Can potentially detect new or zero-day attacks but might generate false positives if the baseline isn't precisely established.

Advantages of Anomaly-based IDS

1. Anomaly-based IDS Detection of unexpected threats: Rather than relying on a database of known attack signatures, it can identify new or previously unknown risks by identifying any activity that deviates from a taught "normal" baseline.
2. Anomaly-Based IDS Shows Adaptability in a such a way it can Adjusts to changes in network behavior over time, keeping it effective even if new applications or legitimate traffic patterns arise.

Disadvantages of Anomaly-based IDS

1. Anomaly-Based IDS Shows High false positive rate: Legitimate but uncommon activity may be wrongly identified as suspicious, resulting in unnecessary warnings.
2. Anomaly-Based IDS have Learning period required in that It requires an initial "learning" phase to determine what constitutes regular network activity. During this time, it may be less effective at detecting irregularities.

Machine learning-based IDS

Utilizes machine learning methods to assess and detect threats. Can be applied in several forms of IDS, such as anomaly-based or behavior-based systems.

Advantages of Machine Learning Based IDS :

a.Machine learning-based IDS Detects novel attacks: Unlike signature-based systems, machine-based IDS can detect previously unknown threats by learning normal network behavior and marking deviations as anomalies.

b.Machine learning-based IDS Adapts to changes: These systems are more flexible than static signature-based systems because they can respond to changes in network traffic, such as upgrades or new apps.

Machine Learning Algorithms :

SVM (Support Vector Machine)

A machine-learning algorithm used for classification and regression.
It works by finding the best separating boundary (hyperplane) between classes. Finds the maximally separating hyperplane. Works well for high-dimensional data. Uses support vectors (important data points near the boundary). Can classify non-linear data using kernel tricks (e.g., RBF, polynomial).Known for high accuracy and robustness.

CNN (Convolutional Neural Network)

A deep learning model mainly used for image, video, and pattern recognition.
It works by using convolutional layers to automatically extract features like edges, shapes, and textures.Learns features automatically (no manual feature engineering).Uses convolutions, pooling, and fully connected layers. Excellent for image classification, object detection, and signal analysis.

Random Forest / Multiple Decision Tree

A machine-learning algorithm that uses many decision trees and combines their outputs to make a more accurate and stable prediction. Ensemble of multiple decision trees. Reduces overfitting. Works well for classification and regression. Robust, accurate, and handles noisy data well.

Clustering

A machine learning technique used to group similar data points together without using labels (unsupervised learning).Finds natural patterns or groups in data. No predefined categories . Common algorithms: k-means, Hierarchical clustering, DBSCAN. Used for anomaly detection, customer segmentation, pattern discovery.

Relative Work Of KDD Dataset

Knowledge Discovery in Databases (KDD) refers to the entire process of extracting useful information from huge datasets. It begins with the selection of relevant data, then moves on to preprocessing to clean and organize it, transformation to prepare it for analysis, data mining to uncover patterns and relationships, and finally evaluation and interpretation of results, all of which result in valuable knowledge or insights. KDD is frequently used in areas such as machine learning, pattern recognition, statistics, artificial intelligence, and data visualization.

This section contains related efforts on leveraging the KDD dataset to develop machine learning techniques. It also gives a quick review of the major machine learning techniques and demonstrates how the KDD dataset is particularly beneficial for analyzing and testing various types of machine learning algorithms. The classifier selection methodology presented by [2] the authors conducted a thorough review of intrusion detection systems and KDD datasets. They retrieved 49596 instances from the KDD dataset to train multiple machine learning methods, including Naive Baye's and multi-layer perceptrons. The authors were successful in proposing two models for detecting various sorts of intrusions in KDD dataset [1].

The authors focus on identifying the most significant qualities to construct IDS with a high accuracy rate and minimal calculation time [4].

Recent DOS/DDOS Attacks

2016 Mirai botnet attack on Dyn. The Mirai botnet, which was made up of compromised Internet of Things devices like IP cameras, searched the internet for susceptible devices and transformed them into a zombie army.

The botnet then launched a massive and coordinated Distributed Denial of Service (DDoS) attack against Dyn, a major DNS service provider. The attack overwhelmed Dyn's servers, making it impossible for users to access many popular websites and online services for hours. According to Investopedia and CovertSwarm, the several impacted services included Amazon, Netflix, Spotify, Twitter, Reddit, and PayPal [Sourse : Internet].

In February 2020, Amazon Web Services (AWS) mitigated one of the largest DDoS attacks ever recorded. Attackers launched the attack by taking advantage of the lax security of third-party networks linked to AWS. They were able to magnify the attack volume by 50 to 70 times every compromised "zombie" client. Although no long-term damage was recorded, the attack's peak incoming traffic rate reached an astounding 2.3 Tbps, affecting AWS for three days. Another of the largest attacks ever recorded, in 2021 is peaked at 3.47 Tbps [3].

KDD Dataset Analysis And Processing :

KDD Knowledge Discovery in Databases (KDD) : KDD refers to the complete process of Extracting knowledge from large datasets. The KDD Process is Iterative , Reliable, and Accurate involving repeated Knowledge Extraction .The Whole Process Consist of Following Steps:

1. Processing of Data
2. Data Mining
3. Pattern Evaluation
4. Knowledge Presentation

KDD Process

KDD is a Process of extracting useful Knowledge from large Volume Of Dataset. NSL-KDD Dataset is an improvement Over KDD-99 Dataset . It is beneficial for cyber security research Particularly, For Building and Evaluating Intrusion Detection System (IDS).NSL-KDD Dataset useful to classify network traffic as normal or Specific type of Attack (Ex. Denial of Service (DOS), Probing, Remote to Local (R2L), User to Root (U2R).NSL-KDD Dataset is More Advantageous than KDD – 99 Dataset [5].KDD Dataset Analysis And Processing KDD Dataset have a various intrusion Behavior at Same time it is used in Several Areas Such as For Testing and Evaluation of Intrusion Detection Algorithms. The First KDD Dataset Was Published in 1999 By MIT Licons labs at University of California [13]. 6 KDD Dataset Includes 4898431 Instances with 41 Attributes . KDD Dataset Imported to the SQL Server 2008 to implement various Statistical measurements values .Examples are Distribution of Instances Records , Attack Types and Accurate Ratio.

Table 1: Showing KDD Attacks

Attack Class	Attack Types
DoS	Black , Lane, Neptune, Pod, Smurf, Teardrop, Apache 2 , Worm (10), Udstorm, Processtable
U2R	Buffer_Overflow , LoadModule , Rootkit , Perl , SqlAttack , Xterm ,Ps(7)
R2L	Guess_Password, FTP_Write, Imap, Phf, Multihop, Spy, Xlock, SendMail, Named(16)
Probe	Satan, Ipsweep, Nmap, Portsweep,, Mscan, Saint(6)

Conclusion

In this paper, we explored various types of Intrusion Detection Systems (IDS) and examined how different machine learning techniques—including SVM, CNN, Random Forest, and clustering—can be applied to detect and prevent cyber threats. The study highlights the potential of AI-driven approaches in enhancing the accuracy, efficiency, and adaptability of IDS, while reducing false positives. Using the KDD-99 dataset as a benchmark, it is evident that the choice of machine learning model significantly impacts detection performance. Overall, the integration of AI and advanced learning algorithms provides a promising direction for developing more intelligent and robust IDS solutions to address evolving cyber security challenges.

References

[1]Nguyen, Huy Anh, and Deokjai Choi. "Application of data mining to network intrusion detection: classifier selection model." *Asia-Pacific Network Operations and Management Symposium*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[2] Almseidin, Mohammad, et al. "Evaluation of machine learning algorithms for intrusion detection system." *2017 IEEE 15th international symposium on intelligent systems and informatics (SISY)*. IEEE, 2017.

[3] Kessler, Gary C. "Defenses against distributed denial of service attacks." *SANS Institute* 2002 (2000).

[4] W. Alsharafat, "Applying artificial neural network and extended classifier system for network intrusion detection." International Arab Journal of Information Technology (IAJIT), vol. 10, no. 3, 2013.

[5] Srikanth Kavuri. (2022). AI-Driven Automation Techniques for Enhanced Software Testing Efficiency. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 456–466.

[6] M. K. Lahre, M. T. Dhar, D. Suresh, K. Kashyap, and P. Agrawal, "Analyze different approaches for ids using kdd 99 data set," International Journal on Recent and Innovation Trends in Computing and Communication, vol. 1, no. 8, pp. 645–651, 2013.

[7] F. Haddadi, S. Khanchi, M. Shetabi, and V. Derhami, "Intrusion detection and attack classification using feed-forward neural network," in Computer and Network Technology (ICCNT), 2010 Second International Conference on. IEEE, 2010, pp. 262–266.

[8] Z. Zhang, J. Li, C. Manikopoulos, J. Jorgenson, and J. Ucles, "Hide: a hierarchical network intrusion detection system using statistical preprocessing and neural network classification," in Proc. IEEE Workshop on Information Assurance and Security, 2001, pp. 85–90.

[9] W. Alsharafat, "Applying artificial neural network and extended classifier system for network intrusion detection." International Arab Journal of Information Technology (IAJIT), vol. 10, no. 3, 2013.

[10] N. Bhargava, G. Sharma, R. Bhargava, and M. Mathuria, "Decision tree analysis on j48 algorithm for data mining," Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 6, 2013.

[11] C. Fleizach and S. Fukushima, "A naive bayes classifier on 1998 kdd cup," 1998.

[12] M. Alkasassbeh, G. Al-Naymat, A. B. Hassanat, and M. Almseidin, "Detecting distributed denial of service attacks using data mining techniques," International Journal of Advanced Computer Science & Applications, vol. 1, no. 7, pp. 436–445.

[13] S. D. Bay, "The uci kdd archive [http://kdd. ics. uci. edu]. irvine, ca: University of california," Department of Information and Computer Science, vol. 404, p. 405, 1999.

[14] M. Al-Kasassbeh, "Network intrusion detection with wiener filter-based agent," World Appl. Sci. J, vol. 13, no. 11, pp. 2372–2384, 2011.

[15] S. K. Pal and S. Mitra, "Multilayer perceptron, fuzzy sets, and classification," IEEE Transactions on neural networks, vol. 3, no. 5, pp. 683–697, 1992.

[16] A. Cutler and G. Zhao, "Pert-perfect random tree ensembles," Computing Science and Statistics, vol. 33, pp. 490–497, 2001.

[17] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5– 32, 2001. [18] J. R. Quinlan, C4. 5: programs for machine learning. Elsevier, 2014.

[18] M. S. Bhullar and A. Kaur, "Use of data mining in education sector," in Proceedings of the World Congress on Engineering and Computer Science, vol. 1, 2012, pp. 24–26.

[19] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," Machine learning, vol. 29, no. 2-3, pp. 131–163, 1997.

[20] R. Kohavi and D. Sommerfield, "Targeting business users with decision table classifiers." in KDD, 1998, pp. 249–253.

[21] P. Aditi and G. Hitesh, "A new approach of intrusion detection system using clustering, classification and decision table," 2013. [23] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," ACM SIGKDD explorations newsletter, vol. 11, no. 1, pp. 10–18, 2009.