ISSN 2063-5346

USING MACHINE LEARNING METHODS IN INTRUSION DETECTION SYSTEM

Nausheen Fatima[1], Shaik Gayazuddin[2], Gali Siddardha Reddy[3], Akula Tejesh[4], Mohammad Faiz[5*], Ramandeep Sandhu[6]

Article History: Received: 01.02.2023 Revised: 07.03.2023 Accepted: 10.04.2023

Abstract

Intrusion detection is very important now a days to ensure the safety of the computer networks. As, the cyber-attacks increasing these days traditional intrusion detection system which uses rule-based methods has failed to perform well. Machine learning assisted in providing solutions to these attacks and threats. Machine Learning observes patterns and learn from it to detect anomalies in the network traffic and help us to avoid those attacks. The intrusion detection using machine learning uses techniques like supervised, unsupervised, and deep learning to build an intrusion detection system that helps us to find out the attacks and threats in the network traffic. In this paper I have used a few machine learning algorithms like Decision-Tree, Random-Forest, Naïve-Bayes, Support-Vector-Machine, Logistic-Regression to classify the attacks in the dataset KDD CUP and test their accuracies to find out the train and test accuracies of the methods. Overall, the paper shows that the machine learning based intrusion detection system improves the network security.

[1,2,3,4,5*,6]School of Computer Science & Engineering Lovely Professional University, Phagwara, Punjab, India

[1]nausheen.28838@lpu.co.in, [2]ugandhar2412@gmail.com, [3]siddustuart143@gmail.com, [4]tejeshchintu502@gmail.com, [5*]faiz.techno20@gmail.com, [6]ramandeepsandhu887@gmail.com

DOI:10.31838/ecb/2023.12.s1-B.210

1. Introducton:

Because modern networks are becoming more complex, traditional methods of intrusion detection are frequently insufficient. Machine learning algorithms provide automatic, precise, and adaptable detection capabilities, making them a promising strategy for enhancing intrusion detection.

Algorithms used by machine learning-based intrusion detection systems can identify trends and anomalies in network traffic data. These algorithms can be honed using past data to spot patterns that point to criminal activity and alert users to potential hazards immediately.[11] The following **figure1** shows the dividing the intrusion detection system.

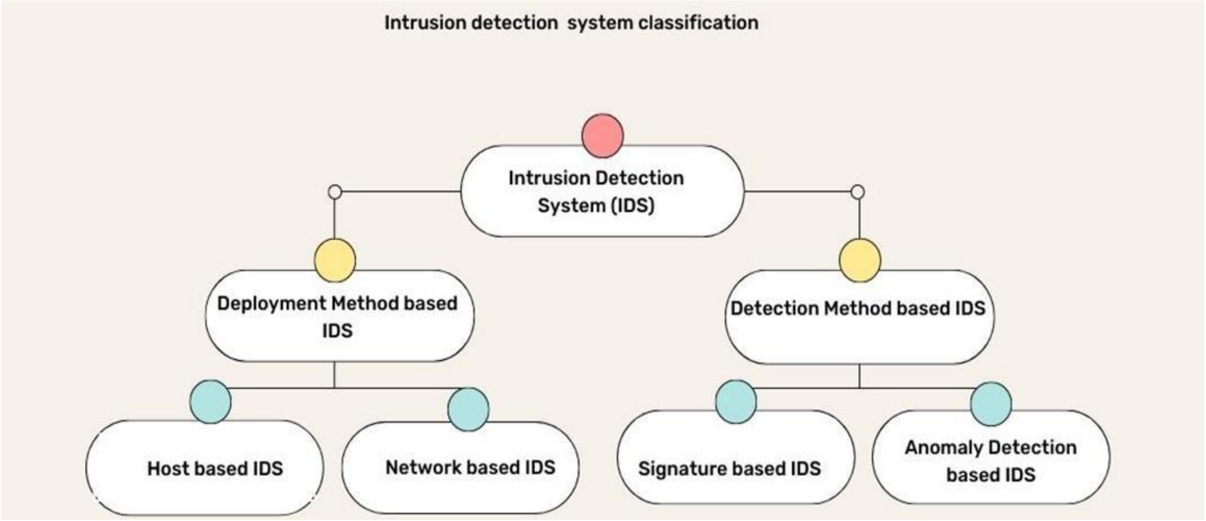

Intrusion detection system classification
Intrusion Detection System (IDS)
Deployment Method based IDS
Detection Method based IDS
Host based IDS
Network based IDS
Signature based IDS
Anomaly Detection based IDS

Figure 1 shows the classification of intrusion detection system

A. Deployment based IDS:

Deployment-based IDS methods are approaches that are used to deploy an IDS in a system or networkenvironment. Here are some common deployment-based IDS methods:

- **Network-based IDS (NIDS):** This method deploys an IDS on a network to monitor and analyze network traffic for potential security breaches. NIDS monitors network traffic, looking for patterns or signatures that indicate a potential attack. This paper is based on the Network-based IDS as shown in **figure 2**.[10]

- **Host-based IDS (HIDS):** This method deploys an IDS on a specific host or server to monitor and analyze activity on that system. HIDS can detect a variety of attacks, including malware infections, unauthorized access attempts, and changes to critical system files.[10]

B. Detection based IDS :

A detection-based IDS recognizes intrusion by processing networks traffic or system events and comparing them to known signatures or patterns of malicious activity.

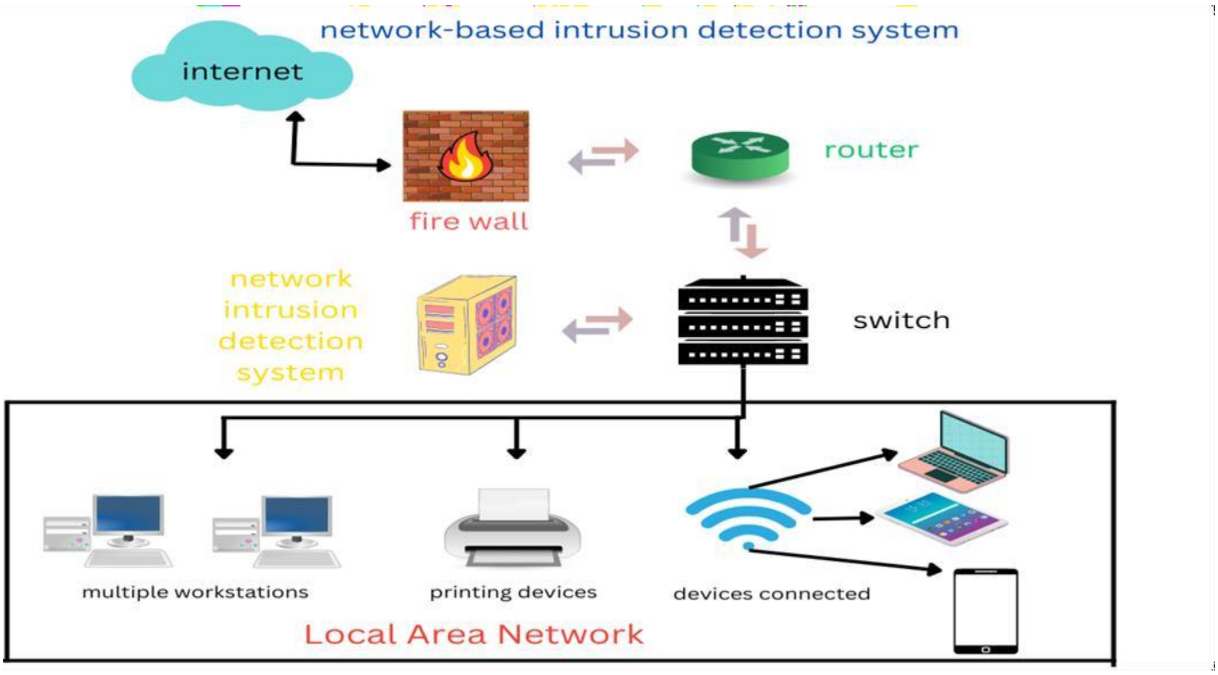


Figure 2 the network-based intrusion detection system

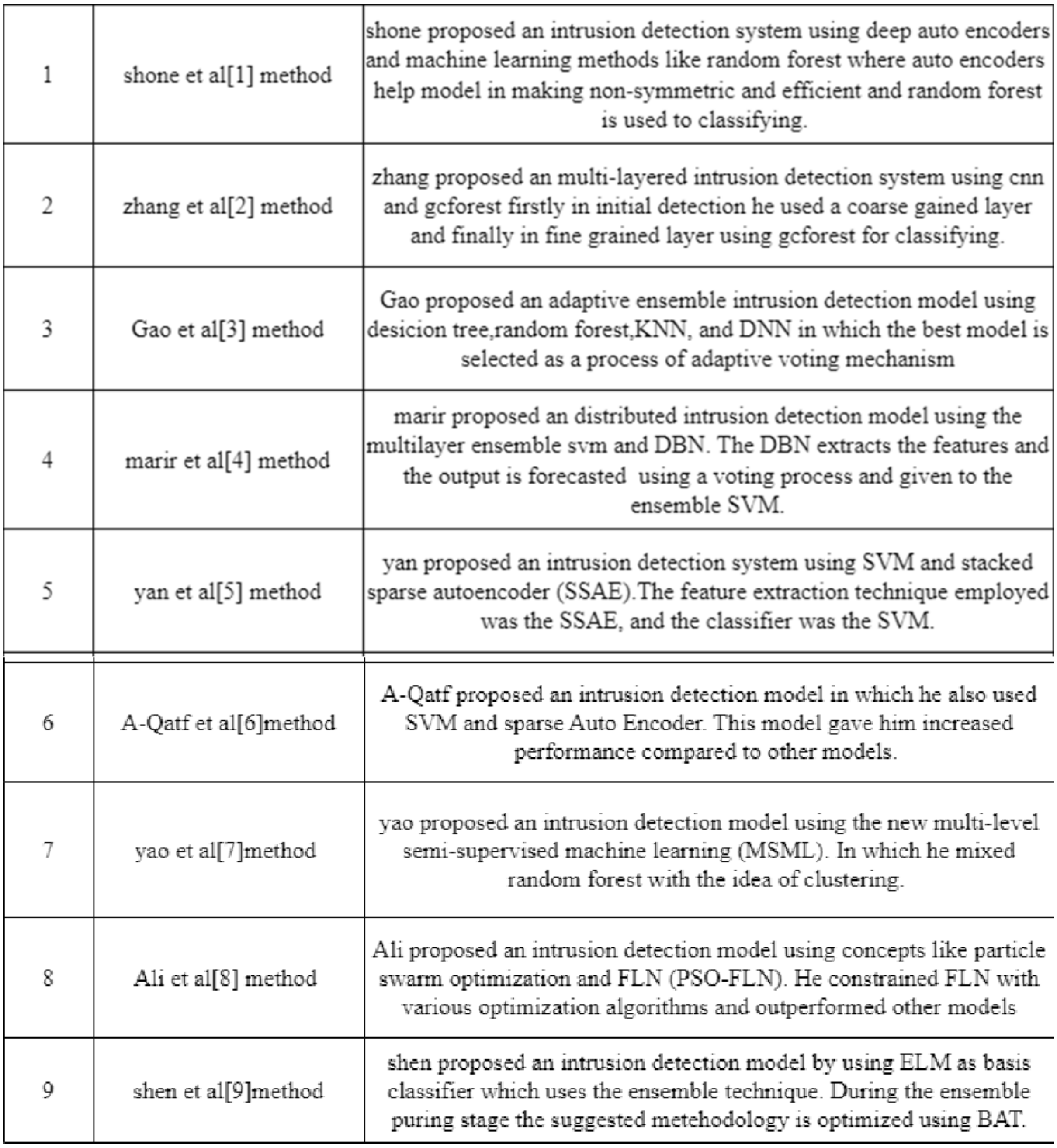
There are several methods that detection-based IDSs can use to identify intrusions:

- **Signature-based detection:** This method generally build upon the database of a known attack signatures to find out traffic which is malicious. [10]
- **Anomaly-based detection:** This method uses machine learning algorithms to analyze normal networktraffic patterns and identify any deviations from the norm. An alert is generated if traffic patterns are deemed to be suspicious.[10]

2. LITERATURE REVIEW

As the recent increase of cyber attacks has led the research on this intrusion detection system more [11-13]. The table 1 shows the research methods of intrusion detection system which are the recent methods for building an intrusion detection model.

Table 1 various methods used for the intrusion detection system



1	shone et al[1] method	shone proposed an intrusion detection system using deep auto encoders and machine learning methods like random forest where auto encoders help model in making non-symmetric and efficient and random forest is used to classifying.
2	zhang et al[2] method	zhang proposed an multi-layered intrusion detection system using cnn and gcforest firstly in initial detection he used a coarse gained layer and finally in fine grained layer using gcforest for classifying.
3	Gao et al[3] method	Gao proposed an adaptive ensemble intrusion detection model using desicion tree,random forest,KNN, and DNN in which the best model is selected as a process of adaptive voting mechanism
4	marir et al[4] method	marir proposed an distributed intrusion detection model using the multilayer ensemble svm and DBN. The DBN extracts the features and the output is forecasted using a voting process and given to the ensemble SVM.
5	yan et al[5] method	yan proposed an intrusion detection system using SVM and stacked sparse autoencoder (SSAE).The feature extraction technique employed was the SSAE, and the classifier was the SVM.
6	A-Qatf et al[6]method	A-Qatf proposed an intrusion detection model in which he also used SVM and sparse Auto Encoder. This model gave him increased performance compared to other models.
7	yao et al[7]method	yao proposed an intrusion detection model using the new multi-level semi-supervised machine learning (MSML). In which he mixed random forest with the idea of clustering.
8	Ali et al[8] method	Ali proposed an intrusion detection model using concepts like particle swarm optimization and FLN (PSO-FLN). He constrained FLN with various optimization algorithms and outperformed other models
9	shen et al[9]method	shen proposed an intrusion detection model by using ELM as basis classifier which uses the ensemble technique. During the ensemble puring stage the suggested metehodology is optimized using BAT.

Overall, these research papers show the potential of machine learning to improve intrusion detection efficiency. Machine learning-based intrusion detection systems have enormous promise for enhancing cybersecurity, but there are still issues that need to be resolved.

3. Methodology and simulation:

Intrusion Detection System actually means a software which is used to detect the attacks and threats in the network traffic. In this intrusion detection model. It observes and learns to say the difference between the attacks and normal which means it creates a classifier in its model which classifies in to attacks and normal [14-16]. Where the attacks in the data set I used are denial-of-service, probing, user-to-root (U2R), remote-to-local9(R2L). The dataset which I have used is kddcup.data-10-percent. Which is 10 percent of the original dataset Kddcup 1999 dataset. I have used various machine learning algorithms like Decision-Tree, Naive-Bayes, Random-Forest, Support-Vector-Machine, Logistic-Regression and used classification methods for each algorithms and compared the results. The steps are as follows [17][18].

- Data Preprocessing: first imported necessary libraries like pandas, numpy, matplotlib, seaborn, os, time etc. and reading the dataset to find missing features and checked correlation on the features to avoid the unnecessary features [19][20].
- Modelling: In this used imported the libraries like sklearn and split the dataset and applied various algorithms like mentioned above and trained and tested this models and got theaccuracies and time of test and train of the models as shown in the following figures:[21][22].

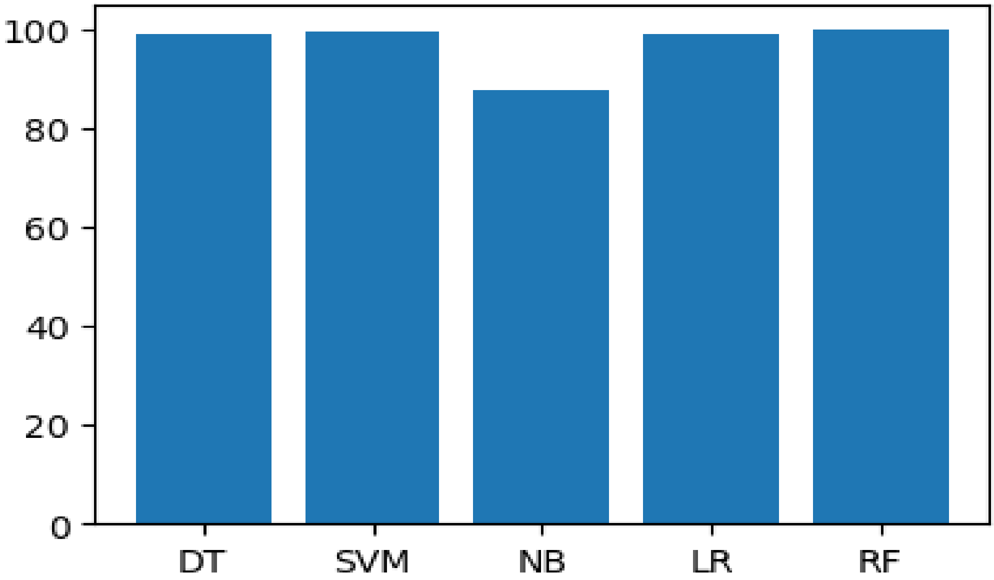


Figure 3 The train accuracy of each model

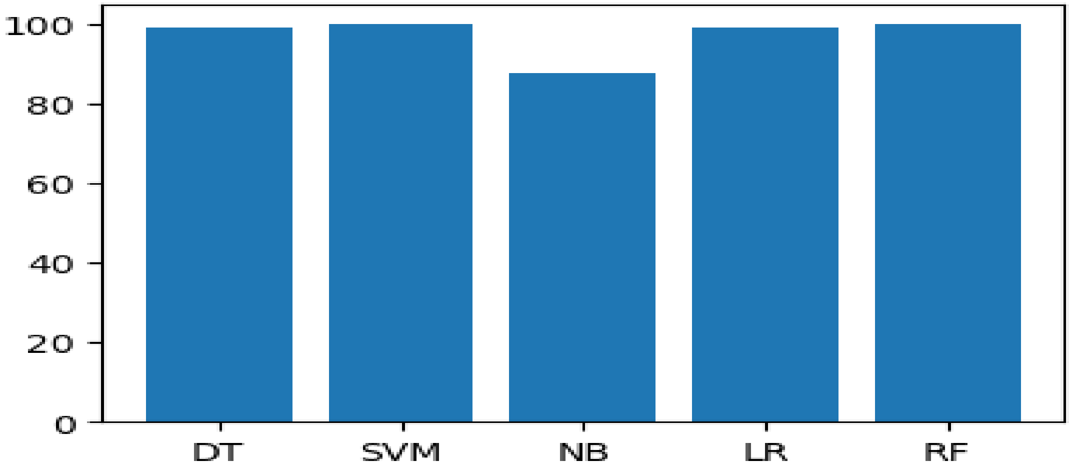


Figure 4 The test accuracy of each model

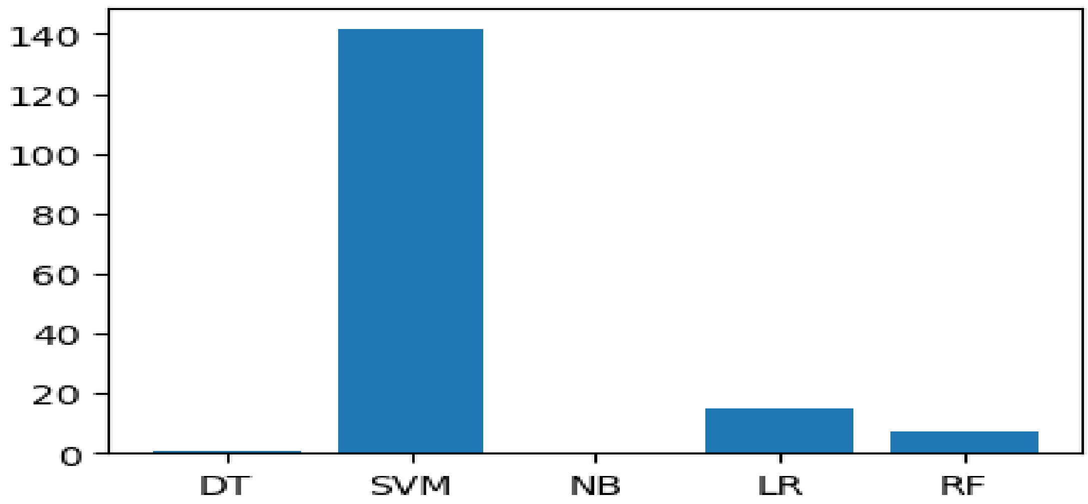


Figure 5 shows the train time of each model

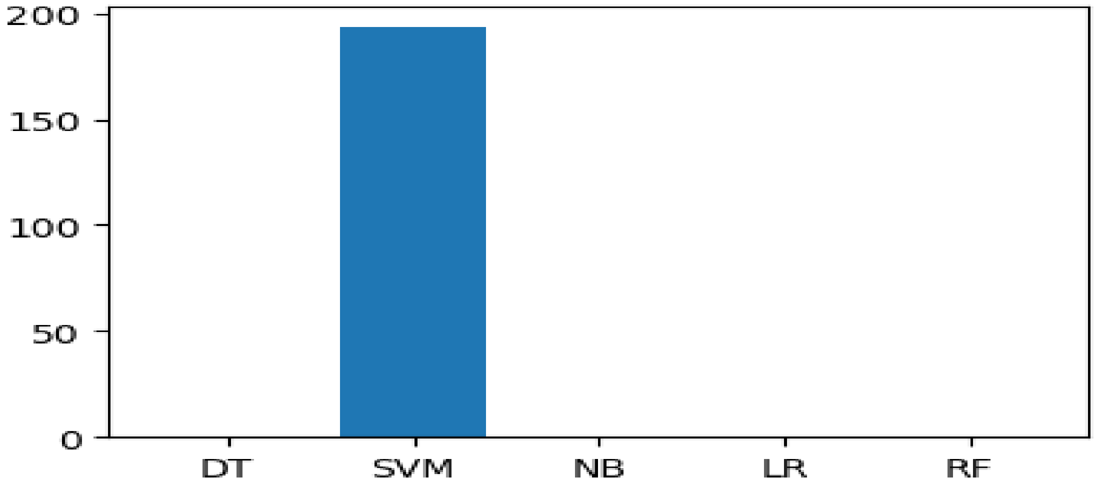


Figure 6 shows the test time of each model

4. Advantages

Intrusion Detection Systems (IDS) have a number of benefits, such as [23][24].

- **Early detection of attacks:** IDS is capable of real-time or nearly real-time detection of potential attacks or security lapses. This enables prompt actions to avert or lessen potential harm.
- **Reduced downtime:** IDS can swiftly identify and contain security issues, minimizing downtime and the consequences of security breaches.
- **Compliance with regulations:** IDS can assist organizations in complying with these rules by enabling real-time monitoring and reporting of security events. Many industries are subject to severe security legislation and standards [25][26].
- **Cost-effectiveness:** Compared to hiring more security professionals or investing in new security hardware, implementing an IDS may be more cost-effective [27][28].
- **Enhanced situational awareness:** IDS can offer a complete picture of a company's security posture, assisting security teams in locating weaknesses and potential attack points [29][30].

- **Improved incident response:** By giving organizations specific information about the type and breadth of the attack, IDS can assist organizations in responding to security problems more quickly and effectively [31][32][33].

5. Disadvantages

In order to detect and notify administrators of potential security lapses and attacks on computer networks or systems, intrusion detection systems (IDS) are used. IDS, like any technology, has its drawbacks as well, such as:

- **False negatives:** Some attacks, such as those that employ cutting-edge tactics or target particular vulnerabilities that the IDS is not built to detect, may go undetected by IDS.
- **Over-reliance on signatures**: Some IDS place a significant emphasis on signature-based detection, whichmeans they search for particular patterns of known harmful behaviour. This may be effective against well- known dangers, but attackers using original or specialized attack techniques can readily get around it.
- **Maintenance and configuration:** IDS need periodic upkeep and configuration to stay functional, which can take a lot of time and need specialized knowledge.
- **Cost:** IDS implementation and upkeep can be costly, particularly for bigger networks or organizations.
- **Privacy concerns:** Some employees or users may find IDS obtrusive and worry about their privacy and data collection.
- **Complex setup and management:** Effective implementation of an IDS can be challenging for small or resource-constrained organizations because setting it up and administering it can be complicated and require specialized skills.

6. Conclusion and Future scope

From the above methods we conclude that decision-tree has highest accuracy and low training and testing time. Therefore, Decision-Tree method best performs on the kddcup dataset. I have used machine learning algorithms like decision-tree, support-vector-machine, random-forest, logistic-regression, naïve-bayes methods to classify in to attacks and normal in which decision-tree method performed well. ML algorithms can be trained on large datasets of known attacks, and can learn to identify previously unknown attacks based on patterns and anomalies in network traffic. As ML algorithms improve, the accuracy of IDS will also improve. Anomaly detection using ML can be applied to identify unusual behavior or patterns that deviate from normal network traffic. This can be especially useful for detecting zero-day attacks or previously unknown threats. Overall, the future scope for IDS using ML is promising, and as ML algorithms continue to improve, IDS will become even more effective in detecting and preventing cyber-attacks.

REFERENCE

[1] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, “A Deep Learning Approach to Network Intrusion Detection,” vol. 2, no. 1, pp. 41–50, 2018.

[2] L. Han and J. Lin, “A Multiple-Layer Representation Learning Model for Network-Based Attack Detection,” IEEE Access, vol. 7, pp. 91992–92008, 2019, doi: 10.1109/ACCESS.2019.2927465.

[3] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, “An Adaptive Ensemble Machine Learning Model for Intrusion Detection,” IEEE Access, vol. 7, pp. 82512–82521, 2019, doi: 10.1109/ACCESS.2019.2923640.

[4] N. Marir, H. Wang, G. Feng, and

B. Li, “Distributed Abnormal Behavior Detection Approach based on Deep Belief Network and Ensemble SVM using Spark,” vol. 3536, no. c, pp. 1–15, 2018, doi: 10.1109/ACCESS.2018.2875045.

[5] B. Yan and G. Han, “Effective feature extraction via stacked sparse autoencoder to improve intrusion detection system,” IEEE Access, vol. PP, no. c, p. 1, 2018, doi: 10.1109/ACCESS.2018.2858277.

[6] M. Al-qatf, M. Alhabib, and K. Al- sabahi, “Deep Learning Approach Combining Sparse Autoen- coder with SVM for Network Intrusion Detection,” IEEE Access, vol. PP, no. c, p. 1, 2018, doi: 10.1109/ACCESS.2018.2869577.

[7] H. Yao, D. Fu, P. Zhang, M. Li, and Y. Liu, “MSML : A Novel Multi-level Semi-supervised Machine Learning Framework for Intrusion Detection System,” IEEE Internet Things J., vol. PP, no. XX, p. 1, 2018, doi: 10.1109/JIOT.2018.2873125.

[8] M. Hasan, B. Abbas, D. Al, and M. A. Binti, “A new intrusion detection system based on Fast Learning Network and Particle swarm optimization,” vol. XX, no. c, 2018, doi: 10.1109/ACCESS.2018.2820092.

[9] Y. A. S. Hen, K. A. Z. Heng, C. W. U. Hunhua, and M. I. Z. Hang, “An Ensemble Method based on Selection Using Bat Algorithm for Intrusion Detection,” no. January, 2018, doi: 10.1093/comjnl/bxx101/4582946.

[10] Z. Ahmad, “Network intrusion detection system : A systematic study of machine learning and deep learning approaches,” no. May 2020, pp. 1–29, 2021, doi: 10.1002/ett.4150.

[11] S. H. Kok, A. Abdullah, N. Z. Jhanjhi, and M. Supramaniam, “A Review of Intrusion Detection System using Machine Learning Approach,” vol. 12, no. 1, pp. 8–15, 2019.

[12] Mall, P. K., Narayan, V., Pramanik, S., Srivastava, S., Faiz, M., Sriramulu, S., & Kumar, M. N. (2023). FuzzyNet- Based Modelling Smart Traffic System in Smart Cities Using Deep Learning Models. In S. Pramanik & K. Sagayam (Eds.), Handbook of Research on Data- Driven Mathematical Modeling in Smart Cities (pp. 76-95). IGI Global. https://doi.org/10.4018/978-1-6684-6408-3.ch005

[13] Faiz, M., Fatima, N., Sandhu, R., Kaur, M., & Narayan, V. (2023). IMPROVED HOMOMORPHIC ENCRYPTION FOR SECURITY IN CLOUD USING PARTICLE SWARM OPTIMIZATION. Journal of Pharmaceutical Negative Results, 2996-3006.

[14] Faiz, M., Daniel, A.K. A multi-criteria cloud selection model based on fuzzy logic technique for QoS. Int J Syst Assur Eng Manag (2022). https://doi.org/10.1007/s13198-022-01723-0

[15] Faiz, M., & Daniel, A. K. (2022). Threats and challenges for security measures on the internet of things. Law, State and Telecommunications Review, 14(1), 71-97.

[16] Choudhary, S., Narayan, V., Faiz, M., & Pramanik, S. (2022). Fuzzy approach- based stable energy-efficient AODV routing protocol in mobile ad hoc networks. In Software Defined Networking for Ad Hoc Networks (pp. 125-139). Cham: Springer International Publishing.

[17] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, “A Deep Learning Approach to Network Intrusion Detection,” vol. 2,

no. 1, pp. 41–50, 2018.

[18] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, "An Adaptive Ensemble Machine Learning Model for Intrusion Detection," IEEE Access, vol. 7, pp. 82512–82521, 2019, doi: 10.1109/ACCESS.2019.2923640.

[19] N. Marir, H. Wang, G. Feng, and B. Li, "Distributed Abnormal Behavior Detection Approach based on Deep Belief Network and Ensemble SVM using Spark," vol. 3536, no. c, pp. 1–15, 2018, doi: 10.1109/ACCESS.2018.2875045.

[20] B. Yan and G. Han, "Effective feature extraction via stacked sparse autoencoder to improve intrusion detection system," IEEE Access, vol. PP, no. c, p. 1, 2018, doi: 10.1109/ACCESS.2018.2858277.

[21] M. Al-qatf, M. Alhabib, and K. Al-sabahi, "Deep Learning Approach Combining Sparse Autoen- coder with SVM for Network Intrusion Detection," IEEE Access, vol. PP, no. c, p. 1, 2018, doi: 10.1109/ACCESS.2018.2869577.

[22] H. Yao, D. Fu, P. Zhang, M. Li, and Y. Liu, "MSML : A Novel Multi-level Semi-supervised Machine Learning Framework for Intrusion Detection System," IEEE Internet Things J., vol. PP, no. XX, p. 1, 2018, doi: 10.1109/JIOT.2018.2873125.

[23] M. Hasan, B. Abbas, D. Al, and M. A. Binti, "A new intrusion detection system based on Fast Learning Network and Particle swarm optimization," vol. XX, no. c, 2018, doi: 10.1109/ACCESS.2018.2820092.

[24] Y. A. S. Hen, K. A. Z. Heng, C. W. U. Hunhua, and M. I. Z. Hang, "An Ensemble Method based on Selection Using Bat Algorithm for Intrusion Detection," no. January, 2018, doi: 10.1093/comjnl/bxx101/4582946.

[25] Z. Ahmad, "Network intrusion detection system : A systematic study of machine learning and deep learning approaches," no. May 2020, pp. 1–29, 2021, doi: 10.1002/ett.4150.

[26] S. H. Kok, A. Abdullah, N. Z. Jhanjhi, and M. Supramaniam, "A Review of Intrusion Detection System using Machine Learning Approach," vol. 12, no. 1, pp. 8–15, 2019.

[27] Babu, S. Z., et al. "Abridgement of Business Data Drilling with the Natural Selection and Recasting Breakthrough: Drill Data With GA." Authors Profile Tarun Danti Dey is doing Bachelor in LAW from Chittagong Independent University, Bangladesh. Her research discipline is business intelligence, LAW, and Computational thinking. She has done 3 (2020).

[28] Narayan, Vipul, A. K. Daniel, and Pooja Chaturvedi. "E-FEERP: Enhanced Fuzzy based Energy Efficient Routing Protocol for Wireless Sensor Network." Wireless Personal Communications (2023): 1-28.

[29] Paricherla, Mutyalaiah, et al. "Towards Development of Machine Learning Framework for Enhancing Security in Internet of Things." Security and Communication Networks 2022 (2022).

[30] Tyagi, Lalit Kumar, et al. "Energy Efficient Routing Protocol Using Next Cluster Head Selection Process In Two-Level Hierarchy For Wireless Sensor Network." Journal of Pharmaceutical Negative Results (2023): 665-676.

[31] Sawhney, Rahul, et al. "A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease." Decision Analytics Journal 6 (2023): 100169.

[32] Srivastava, Swapnita, et al. "An Ensemble Learning Approach For Chronic Kidney Disease Classification." Journal of Pharmaceutical Negative Results (2022): 2401-2409.

[33] Narayan, Vipul, et al. "Deep Learning Approaches for Human Gait Recognition: A Review." 2023 International Conference on Artificial Intelligence and Smart Communication (AISC). IEEE, 2023.