EasyChair Preprint

№ 8578

Intrusion Detection System: Challenges in Network Security and Machine Learning

Nursyafiqah Abdul Samat

August 3, 2022

Intrusion Detection System: Challenges in Network Security and Machine Learning

Nursyafiqah Abdul Samat[1]

[1]*University Technology MARA, Jalan Ilmu 1/1, Shah Alam, Malaysia*
2021344697@isiswa.uitm.edu.my

Keywords: Intrusion Detection System, Challenges, Machine Learning, Techniques, Disadvantages, Security

Abstract: Intrusion Detection Systems (IDSs) are a type of security management system that can detect and detect attacks aimed at compromising system security elements like accessibility, integrity, and privacy. There is 76% of DDoS attacks were conducted with mixed attacks. In 2020, about 24.02% of attacks employed a single type of technique, with an average of 1.66 attack strategies used per event among those that used multiple methodologies. In this paper, the details of intrusion detection system that include the techniques, methods, and machine learning algorithms are being discussed. There are also advantages and disadvantages that being explain in this paper. Moreover, it leverages its huge attack signature database to monitor the operation of routers, firewalls, important servers, and files, raising an alarm and sending relevant notifications when a breach is detected. Attackers gain access to other people's files by cracking their passwords, resulting in a significant breach in network security. Therefore, developers need to work on more precise, faster, and scalable techniques in order to prevent it from happen.

.

1.0 INTRODUCTION

Malicious software, such as malware, has evolved over time, posing a significant issue to the development of Intrusion Detection Systems (IDS). Malicious attacks have evolved, and the most difficult task is identifying unknown and obfuscated malware, since malware developers employ various evasion tactics for information concealment in order to avoid detection by an IDS [1]. [2] stated that a secure network is known by its hardware and software resistance to various types of attacks.

Based on report by [3], there is 76% of DDoS attacks were conducted with mixed attacks. In 2020, about 24.02 percent of attacks employed a single type of technique, with an average of 1.66 attack strategies used per event among those that used multiple methodologies. It is also stated that proxy server attacks have become the primary attack force in 2020 as it seems simple to procure, cost-effective, and well-executed.

Moreover, security concerns such as zero-day attacks specifically targeting online users have increased. Multiple tools have been designed and employed in various sorts of network attacks because of the rise of several types of attacks. Intrusion detection systems (IDSs) are one of these solutions. As stated by [4], this program can monitor a variety of network systems, as well as cloud computing and information systems. Also, the Intrusion Detection System (IDS) can track and identify attacks aimed at compromising system security elements like accessibility, integrity, and privacy.

Since the network and computer technology are becoming dependent day by day, providing protection and privacy become one of the most difficult challenges facing security management system developers as it should emphasises the importance of secure networks. It is due to protect servers from intruders as intruders have more opportunities to launch malicious attacks when data is being generated from multiple sources [5]. In Intrusion Detection System (IDS) there are three detection method that can be used which are Signature-based, Anomaly-based and Hybrid-based detection. Signatures are the patterns that the IDS detect.

2.0 INTRUSION DETECTION SYSTEM

The process of identifying acts that attempt to threaten a resource's overall privacy and consistency is known as intrusion detection. Even though some of the characteristics could be duplicated and contribute little to the detection process, IDS check the full data characteristics to identify any incursion and exploitation tendencies [6]. As stated by [4], an Intrusion Detection System (IDS)'s job is to keep an eye on a system or network and detect any unusual activity. IDSs are classified as either host-based or network-based. There are few types of IDS that can be separated by the types of Intrusion Detection System and methods used in analyzing the procedures which are Network-based IDS (NIDS), Host-based IDS (HIDS) and Hybrid or mixed IDS (MIDS) that specified by [5].

According to [7] in [4], a packet sniffer is a program that reads raw packets from a section of the local network, that is scanned by a Network-based IDS (NIDS). It can also follow more network purposes in the hopes of detecting dangers that Host-based IDS (HIDS) would miss because it cannot read packet headers or recognize certain types of attacks. NIDSs do not use the host's operating system as an identity source, whereas HIDSs do. Table 2.0 is provided to show the details of the types of Intrusion Detection System (IDS).

Table 2.0: Details of Types of Intrusion Detection System (IDS)

IDS Types	Details
NIDS	It monitors network connectivity in order to detect breaches. It tries to detect illegal, unlawful, and unusual actions only based on network traffic [8].
HIDS	It keeps track of the host's or device's network activity and events. It can identify breaches by examining any changes that occur within hosts in order to uncover unauthorized behavior [9]. This can be accomplished by matching a predetermined pattern to the operating system logs.
MIDS	For more efficient and productive detection of cyberattacks, it merges host-based (HIDS) and network-based (NIDS) detection in a network.

Also, there are three types of detection techniques that can be used on IDS are known as Signature-based IDS or Misuse Detection, Anomaly-based and Hybrid-based [9]. The three basic types of detection techniques utilized in IDS are statistical approaches, data-mining methods, and machine learning (ML). Figure 2.0 review the Intrusion Detection System (IDS).

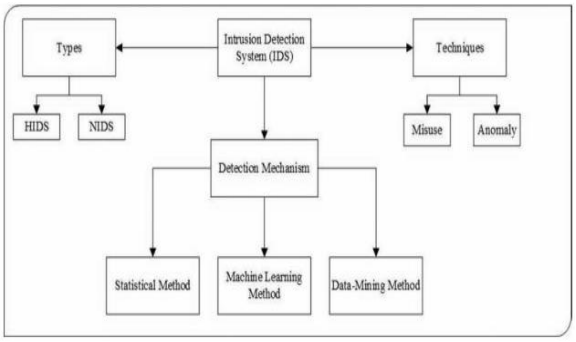

Figure 2.0: Overview of Intrusion Detection System (IDS)

Source: [3]

2.1 DETECTION TECHNIQUES OF INTRUSION DETECTION SYSTEM

Table 2.1 is provided in this section to show the details of detection techniques which are Signature-based known as Misuse Detection, Anomaly-based and Hybrid-based.

Table 2.1: Details of Detection Techniques in Intrusion Detection System (IDS).

Detection Techniques	Details
Signature-based	It refers to the process of detecting network attacks by looking for specific data patterns. Signature is referring to these patterns.
Anomaly-based	This was initially implemented to discover unknown or zero-day threats, which is partially owing to malware's rapid evolution. Machine learning algorithms are used to develop a model, which is then compared to behavior [10].
Hybrid-based	The combination of a signature-based IDS with an anomaly-based IDS.

2.2 INTRUSION DETECTION SYSTEM METHODS

The details of data-mining methods, machine learning (ML) and statistical approaches are discussed in this section. First and foremost, data-mining techniques extract similarities from large amounts of data and aid in decision-making. It is utilized in the field of intrusion detection to learn about invaders' habits and uncover previously unknown patterns of hostile conduct. Clustering is an unsupervised strategy that deals with unlabeled data, while Classification is a supervised machine learning method that requires a preset effort to label data. [11] claimed that these two strategies are the most used strategies in Intrusion Detection, particularly in network data. Similarly, an analysis of ID based on Data Mining Techniques, including the benefits and drawbacks of each methodology, as well as how these might be combined to obtain high-quality Intrusion Detection is presented in [12].

Next, some algorithms, such as Markov, Fuzzy, and Genetic, are the potential statistical approaches. According to [13] and [14], the most probabilistic technique is Markov chain that is based on building a model using normal activity such as audit data. There is a claim in fuzzy logic-based methodologies, an observation is only regarded normal if it falls within a certain range. The best subset is identified using a genetic approach, which boosts classification accuracy by providing the essential inputs to the classification algorithms.

Lastly, machine learning approaches are being used to build IDSs, according to [15]. Machine learning is a type of artificial intelligence approach that can find meaningful information from large datasets automatically. Machine learning's goal is to uncover knowledge and make informed decisions. There are two based on machine learning algorithm which are supervised, unsupervised, and semi-supervised [16]. The common machine learning algorithms used in IDSs are shown in Figure 2.1.

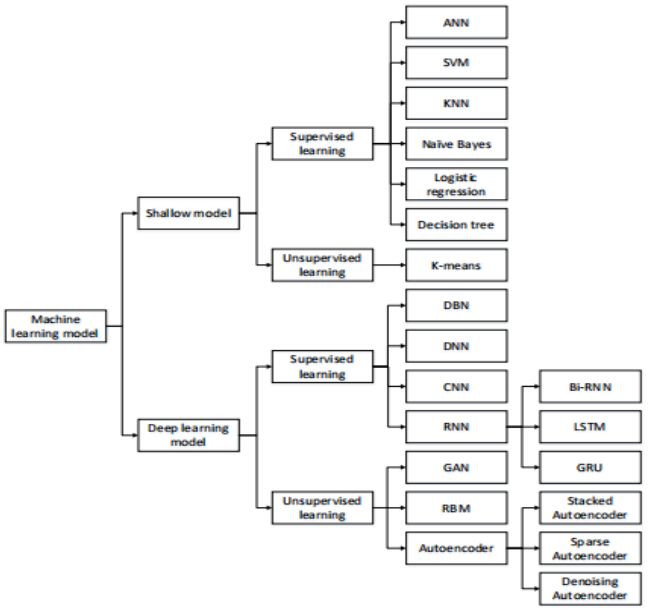

Figure 2.1: Taxonomy of Machine Learning Algorithm

[16] stated that a type of machine learning is supervised learning. In terms of specific classes, it identifies clusters with a high chance density. Data labelling by hand, on the other hand, is both expensive and time-consuming. As a result, supervised learning suffers from a basic bottleneck: a scarcity of labelled data. While unsupervised learning extracts important feature information from data that has not been labelled. When a similarity or differences metric is used, it maximizes intraclass similarities while reducing interclass similarities. Unsupervised approaches typically outperform supervised approaches in terms of detection.

Artificial neural networks (ANN), K-nearest neighbor (KNN), naïve Bayes, support vector machines (SVM), decision trees (DT) and logistic regression (LR) are the supervised learning algorithm and there is only one algorithm which is K-means algorithm when using clustering in unsupervised learning.

2.2.1 ADVANTAGES AND DISADVANTAGES OF EACH SHALLOW ALGORITHM

Each shallow algorithm has its own advantages and disadvantages. The advantages and disadvantages are presented in table 2.2.

Supervised **Algorithm**	**Advantages**	**Disadvantages**
Artificial Neural Network (ANN)	-Capacity to work with nonlinear data -excellent adapting skills	-Training process is time demanding -It is prone to overfitting -It is prone to becoming stuck in a local optimum.
K-Nearest Neighbour (K-NN)	-Implement on large amounts of data -Appropriate for nonlinear data -Train as quickly and efficiently as possible -Noise-resistant	-Low reliability on the minority class -Lengthy testing time -K parameter sensitivity
Naïve Bayes	-Noise-resistant -Capable to learn progressively	-Do not work well on attribute-related data
Support Vector Machines (SVM)	-Small train set provides relevant data -Excellent generating ability	-Do not work well in tasks requiring large amounts of data or many classifications -Kernel function parameters are complex.
Decision Tree (DT)	-Features are automatically selected -Robust analysis	-Neglect the data correlation -Classification result tends to the majority class.
Logistic Regression (LR)	-Simple and quick to learn -Automatically scales features	-Nonlinear data does not work properly -Overfitting is a common occurrence.
Unsupervised Algorithm	**Advantages**	**Disadvantages**
Clustering (K-Means)	-Simple and quick to learn -Scalability is excellent. -Can be used for large amounts of data	-Nonconvex data is difficult to analyze -Initialization is a source of sensitivity -K parameter sensitivity

3.0 INTRUSION DETECTION SYSTEM CHALLENGES

In today's network environment, [17] stated that network security has proven to be a more intricate and hard issue. Attackers from the same company gain access to other people's files by cracking their passwords, resulting in a significant breach in network security. Developers need to work on more precise, faster, and scalable techniques for this. By

2020, the number of linked devices is predicted to approach 26 billion, thanks to the entrance of the "IoT" and Big Data age [18]. There are few challenges that effect the issues which are false alarm, unbalanced dataset, low detection rate and response time [19] as shown in the figure below.

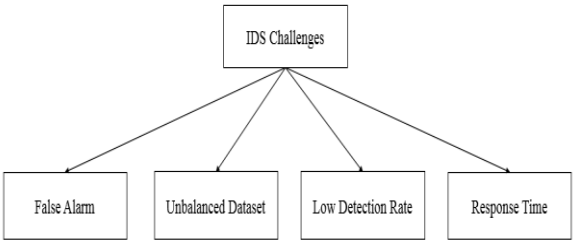

Figure 2.2: IDS Challenges

IDS, according to some experts, should be divided into five categories. Because of the similarities in the power of the techniques, such a classification can cause confusion. Signature or rule-based IDS are typically associated with high false positive rates (FP) alarm rates and unable to identify innovative threats [20]. Systems that rely on stateful protocol analysis perform differently depending on how detailed their profiles are defined. Maintaining an up-to-date profile when new protocols emerge is a big difficulty with this method. IDSs are intended to have a high rate of False Positive (FP) detection capability [21].

Due to the observed minimal discrepancies between normal and harmful behaviors, [22] estimates that around 99% of Intrusion Detection warnings do not represent cybersecurity risks. As the situation associated with an anomaly change, the Alert Classification (ALAC) system by [22] can be incrementally improved. The key obstacles of developing a system with such features are that the complexity of developing such features increases when the system is misused, and control is lost. Finally, there is a greater percentage of false alarms [23], and a low detection rate [24,25] that the uneven dataset also has an impact on the model evaluation [21].

4.0 CONCLUSION

In this paper review, the details of intrusion detection system that include the techniques, methods and machine learning algorithms are being discussed. There are a few more advantages of an intrusion detection system, such as the fact that it enables speedy and effective detection of recognized abnormalities with a minimal chance of false alarms by using a signature database. It also analyses various sorts of assaults, detects dangerous content patterns, and assists administrators in tuning, organizing, and implementing effective restrictions. Finally, it leverages its huge attack signature database to monitor the operation of routers, firewalls, important servers, and files, raising an alarm and sending relevant notifications when a breach is detected.

Despite that, this paper also will show the disadvantages of Intrusion Detection System in order to give awareness to any future researchers that might want to explore more. First and foremost, intrusion detection system requires a full-time monitoring and highly skilled staffs for running it. Other than that, the range of price for installation is quite expensive. Not to forget that it might generates false positives and negatives.

5.0 REFERENCES

[1] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," Cybersecurity, vol. 2, no. 1, 2019, doi: 10.1186/s42400-019-0038-7.

[2] Alibaba Clouder, "DDoS Attack Statistics and Trend Report by Alibaba Cloud," 2021. [Online]. Available: https://www.alibabacloud.com/blog/ddos-attack-statistics-and-trend-report-by-alibaba-cloud_597607.

[3] M. Aljanabi, M. A. Ismail, and A. H. Ali, "Intrusion detection systems, issues, challenges, and needs," Int. J. Comput. Intell. Syst., vol. 14, no. 1, pp. 560–571, 2021, doi: 10.2991/ijcis.d.210105.001.

[4] O. Ahmed and W. Abdullah, "A Review of Intrusion Detection Systems," Acad. J. Nawroz Univ., vol. 6, no. 3, pp. 106–111, 2017, doi: 10.25007/ajnu.v6n3a91.

[5] R. Alshamy and M. Ghurab, "A Review of Big Data in Network Intrusion Detection System: Challenges, Approaches, Datasets, and Tools," Int. J. Comput. Sci. Eng., vol. 8, no. 7, pp. 62–75, 2020, [Online]. Available: https://doi.org/10.26438/ijcse/v8i6.115.

[6] W. Lee and S. J. Stolfo, A Framework for Constructing Features and Models for Intrusion Detection Systems, vol. 3, no. 4. 2000.

[7] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho, "A survey of network-based intrusion detection data sets," Comput. Secur., vol. 86, pp. 147–167, 2019, doi: 10.1016/j.cose.2019.06.005.

[8] Y. Zhang, S. Luo, L. Pan, and H. Zhang, "Syscall-BSEM: Behavioral semantics enhancement method of system call sequence for high accurate and robust host intrusion detection," Futur. Gener. Comput. Syst., vol. 125, pp. 112–126, 2021, doi: 10.1016/j.future.2021.06.030.

[9] M. Verkerken, D. Laurens, T. Wauters, B. Volckaert, and F. De Turck, "Towards Model Generalization for Intrusion Detection :," J. Netw. Syst. Manag., vol. 7, 2022, doi: 10.1007/s10922-021-09615-7.

[10] S. Krishnaveni, P. Vigneshwar, S. Kishore, B. Jothi, and S. Sivamohan, "Anomaly-Based Intrusion Detection System Using Support Vector Machine," 2020, [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-15-0199-9_62.

[11] M. Jouad, S. Diouani, H. Houmani, and A. Zaki, "Security challenges in intrusion detection," Proc. 2015 Int. Conf. Cloud Comput. Technol. Appl. CloudTech 2015, no. June, 2015, doi: 10.1109/CloudTech.2015.7337012.

[12] K. Wankhade, S. Patka, and R. Thool, "An overview of intrusion detection based on data mining techniques," Proc. - 2013 Int. Conf. Commun. Syst. Netw. Technol. CSNT 2013, pp. 626–629, 2013, doi: 10.1109/CSNT.2013.134.

[13] C. M. Chen, D. J. Guan, Y. Z. Huang, and Y. H. Ou, "Attack sequence detection in cloud using hidden Markov model," Proc. 2012 7th Asia Jt. Conf. Inf. Secur. AsiaJCIS 2012, no. 1, pp. 100–103, 2012, doi: 10.1109/AsiaJCIS.2012.24.

[14] P. Kumar, Nitin, V. Sehgal, K. Shah, S. S. P. Shukla, and D. S. Chauhan, "A novel approach for security in Cloud Computing using Hidden Markov Model and clustering," Proc. 2011 World Congr. Inf. Commun. Technol. WICT 2011, pp. 810–815, 2011, doi: 10.1109/WICT.2011.6141351.

[15] H. Liu and B. Lang, "Machine learning and deep learning methods for intrusion detection systems: A survey," Appl. Sci., vol. 9, no. 20, 2019, doi: 10.3390/app9204396.

[16] L. Wang and C. A. Alexander, "Machine learning in big data," Int. J. Math. Eng. Manag. Sci., vol. 1, no. 2, pp. 52–61, 2016, doi: 10.33889/ijmems.2016.1.2-006.

[17] M. Sahu and S. K. Das, "A review on Intrusion Detection System and its future," vol. 2, no. 11, pp. 1351–1355, 2013.

[18] D. Reinsel, J. Gantz, and J. Rydning, "Data Age 2025 : Don 't Focus on Big Data; Focus on the Data That's Big," IDC White Pap., no. April, pp. 1–25, 2017, [Online]. Available: https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf%0Ahttps://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf%0Ahttps://www.seagate.com/files/www-content/our.

[19] M. Umer, H. Xiaoli, and S. Abdul, "Big Data Security Analysis in Network Intrusion Detection System," Int. J. Comput. Appl.,

vol. 177, no. 30, pp. 12–18, 2020, doi: 10.5120/ijca2020919759.

[20] H. Liao, C. R. Lin, Y. Lin, and K. Tung, “Journal of Network and Computer Applications Intrusion detection system : A comprehensive review,” J. Netw. Comput. Appl., vol. 36, no. 1, pp. 16–24, 2013, doi: 10.1016/j.jnca.2012.09.004.

[21] J. M. Fossaceca, T. A. Mazzuchi, and S. Sarkani, “MARK-ELM : Application of a Novel Multiple Kernel Learning Framework for Improving the Robustness of Network Intrusion Detection,” Expert Syst. Appl., 2014, doi: 10.1016/j.eswa.2014.12.040.

[22] T. Pietraszek, “LNCS 3224 - Using Adaptive Alert Classification to Reduce False Positives in Intrusion Detection,” pp. 102–124, 2004.

[23] S. Ali, R. Shah, and B. Issac, “machine learning to Snort system,” Futur. Gener. Comput. Syst., 2017, doi: 10.1016/j.future.2017.10.016.

[24] I. Raghav, “Intrusion Detection and Prevention in Cloud Environment : A Systematic Review,” vol. 68, no. 24, pp. 7–11, 2013.

[25] R. Singh, H. Kumar, and R. K. Singla, “An intrusion detection system using network traffic profiling and online sequential extreme learning machine,” Expert Syst. Appl., 2015, doi: 10.1016/j.eswa.2015.07.015.