Universe International Journal of Interdisciplinary Research
(Peer Reviewed Refereed Journal)
UIJIR | ISSN (O) – 2582-6417
MAY 2021 | Vol. 1 Issue 12
www.uijir.com

REVIEW THE INTRUSION DETECTION SYSTEM (IDS) TECHNIQUE USING MACHINE LEARNING

Author's Name: [1]Prof Khushbu Rai, [2]Ms. Geeta Das

Affiliation: *[1]Professor Of Lakshmi Narain College of Technology & Science, Madhya Pradesh, India*

[2]Student, Lakshmi Narain College of Technology & Science, Madhya Pradesh, India

E-Mail: das.milli7049@gmail.com

DOI No. – 08.2020-25662434

Abstract

Machine learning techniques are widely used in the development of intrusion detection systems (IDS) capable of identifying and distinguishing cyberattacks at the network and host levels in a timely and automatic manner. However, since malicious attacks are continually emerging and occurring in large numbers, some issues arise that necessitate a modular solution. Various ransomware datasets are publicly available for the computer security community to review more. However, no recent study has investigated the performance of various machine learning algorithms on a number of publicly available datasets. This survey proposes an IDS taxonomy that categorises and summarises IDS literature focused on machine learning and deep learning, with data objects serving as the primary component. We both accept that this kind of taxonomy scheme is suitable for computer security researchers. The survey first clarifies the description and taxonomy of IDSs. Then, the machine learning techniques, metrics, and benchmark datasets that are often used in intrusion detection systems are addressed.

Keywords*: Intrusion detection system(IDS), Support Vector Machine(SVM)*

INTRODUCTION

Large network traffic sizes, highly unequal data transmission, the difficulty of distinguishing between natural and irregular operation, and the need for constant adaptation to an ever-changing environment are all obstacles that an IDS must face [14]. In general, the challenge is to effectively identify and classify various operations in a computer network. The two most general forms of network behaviour classification techniques are misuse detection and anomaly detection [4]. Misuse recognition techniques use signature matching algorithms to search for known instances of misuse of both network and system behaviour. This approach is effective for detecting previously known threats. . Indeed, new attacks are often ignored, resulting in false negatives. The intrusion detection system can generate warnings, but reacting to each one consumes time and energy, allowing the system to become unreliable. To address this problem, IDS should not begin the removal process as soon as the first symptom is detected, but should instead wait until all of the warnings have been received before making a decision based on their link. Anomaly detection systems concentrate on the development of a normalised model of user interaction. The combination of mathematical and machine learning approaches to check network traffic, system calls, and processes is used to do this. Since the anomaly detection strategy is more effective at identifying novel threats, any deviant behaviour is considered an intruder. Normal behaviour, on the other hand, is poorly characterised and differs over time in a large and dynamic world. This frequently results in a high number of false alarms, also known as false positives. A network-based intrusion detection system (IDS) scans incoming network traffic for patterns that suggest if someone is searching for infected computers on the network. IDS does not respond to any warning because it will take too much time and energy to respond to

each one. Ignoring this fact can result in a self-inflicted denial of service. To resolve this problem, notifications should be aggregated and correlated, yielding less but more articulate and noticeable alarms.

RELATED WORK

Srivastava, A et al[1] Intrusion detection is a quickly growing field. In the past, supervised learning methods were used to detect intrusions in network traffic data. However, not only has the level of traffic increased significantly in recent years, but network risks have grown as well. To recognise these new types of attacks, improved detection techniques are needed. Deep learning algorithms for detecting anomalies in network traffic have been widely analysed by researchers. Fresh datasets have been introduced to the online collections. In this article, we used novel feature reduction-based machine learning algorithms to detect anomalous patterns in a recently presented dataset. A high precision of 86.15 percent was achieved.

S. S. Swarna Sugi et al[2] This article describes how to combat security threats in IoT networks with an intruder detection system for profound Intrusion Detection System(IDS). The attack detection models have Long Short-Terms Memory (LSTM) and K-Nearest Nears (KNN) and measure their performance with the use of time, statistics, geometrical mean and sensibilities for detection. To test the efficacy of the mounted IDS, the Bot-IoT databases are used.

V. Bhatia et al[3] The IDS's role is to monitor the system in advance for any malicious errors. Criteria such as feature collection, LAND attack tracking, traffic filtering, and blocking, among others, are used to compare the above techniques. According to the feature selection procedure, Anomaly Detection is by far the best tool for Intrusion Detection. This article explains the various types and techniques for an intrusion detection system. Using the IDS approach, the paper explores how neural networks and deep learning can be used to identify intruders and prevent malicious errors in the system.

V. V. Kumari et al[4] This paper demonstrates how to build an efficient intrusion detection system (IDS) by combining Active Learning Support Vector Machine (ASVM) and Fuzzy C-Means (FCM) clustering. This algorithm was found to be promising in the NSL KDD bench mark IDS data set.

R. Vijayanand et al[5] To identify threats efficiently, the proposed approach employs many multi-layer deep algorithms organised in a hierarchical order. The efficiency of the proposed architecture is compared to that of simple multilayer deep learning algorithms and hierarchical SVM algorithms with feature selection methods on the standard CICIDS 2017 dataset.

M. S. Koli et al[6] The paper illustrates the use of machine learning techniques, the k-means classification algorithm, and an alternate variant of the supportive vector rating algorithm to dynamically distribute the standard payload of the packet and classify deviations. Our solution reveals that the hybrid algorithm proposed exceeds its precision in detection. the previously used open-source Snort framework.

D. Subramanyam et al[7] In this article, we will look at how to use an intrusion detection system(IDS) to fix a dos, probe-style attack using a classification-based approach. Using the

knowledge discovery package, numerous attacks aid in remembering different data sets. As a consequence, in this article, we will conclude that machine learning techniques can be used to detect and prevent attacks.

PROPOSED METHODOLOGY

Support Vector Machine (SVM) By building nonlinear decision bounds, this process executes regression and classification tasks. Support vector machines can perform a range of different degrees of difficulty classification and regression tasks due to the nature of the function space in which these boundaries are placed. Support Vector Models may be linear, polynomial, RBF, or sigmoid in form and size[14].

Naive Bayes It was a very well popular Bayesian method for performing classification tasks. Naive Bayes models are reliable and simple to use and analyze classification methods since the independent variables are statistically independent. Where the independent space dimensionality (i.e. the number of input variables) is strong, Naive Bayes is a secure solution (a problem known as the curse of dimensionality). For the reasons stated above, Naive Bayes may outperform other more advanced classification methods. Various methods, including ordinary, lognormal, gamma, and Poisson distributions, can be used to model the conditional distributions of the inputs. (15.)

The k-Nearest Neighbour Algorithm calculates the distance between two points. As compared to other statistical methods, k-Nearest Neighbors is a memory-based method that requires no instruction (i.e., no model to fit). It is part of the genre of prototypes. It is based on the premise that the same objects are more likely to be in the same group. KNN projections are based on a number, by a majority vote (for classification tasks), of sample instances used to model the new ones (i.e. unseen) and on the average one (for regression) (hence the word k-nearest neighbors) [13].

The most frequent perpetrators of intrusions are unauthorised apps, also known as attackers. An hacker may use the Internet to obtain remote access to a computer or disable a service. Understanding how to target a system effectively is needed for successful intrusion detection. In general, an attack can be divided into five phases. The five stages are reconnaissance, exploitation, reinforcement, consolidation, and pillage. An intruder can be observed within the first three levels, but if it reaches the fourth or fifth phase, the system is totally compromised. Abusing, subverting, or breaking a service is all words that can be applied to it. Exploiting websites with stolen passwords or dictionary attacks is an example of server exploitation, while SQL injection is an example of subversion assault. Following an unauthorised forced entry into a building, an attacker performs a camouflage procedure and then instals supplementary equipment and services to take advantage of the privileges gained during the reinforcing phase. Using the hacked user account, an attacker tries to gain full server access. Finally, an attacker takes use of the programmes that are available from the open user account. During the consolidation process, an attacker takes complete ownership of the system as well as the installed backdoor, which is used for communication. The final move is pillage, which involves an attacker's malicious acts such as data manipulation, CPU time stealing, and impersonation. Since computers and networks are designed and programmed by humans, both hardware and software flaws are likely. Human errors and bugs can lead to vulnerabilities [8].

Confidentiality, data integrity, and affordability are the fundamental building blocks of information processing. Authenticity and clarity are also essential in terms of information security. In general, attacks against privacy address passive attacks such as eavesdropping, while attacks against integrity address active attacks such as system scanning attacks such as 'Probe,' and attacks against availability address attacks such as denial of service ('DoS') and distributed denial of service ('DDoS') that render network services unavailable to daily users. IDS networks frequently have limited detection capabilities for eavesdropping attacks. A 'probe' attack may occur over a network or locally inside a system. An attack is now described as a pattern of behaviour that has the potential to breach the confidentiality, data integrity, availability, or any other security policy of a resource. The primary purpose of an IDS device is to detect all of these types of threats in order to secure computers and networks from malicious behaviour. This study focuses on the categorization scheme suggested by the DARPA Intrusion Detection Evaluation.

Classification by Detection Methods Misuse detection is another term for signature-based authentication. The basic philosophy of using signatures to represent attack behaviours. The identifying mechanism matches the signatures of samples using a signature index. The most challenging part of designing misuse recognition schemes is producing successful signatures. Misuse detection has a low false alarm rate and offers comprehensive reports on attack modes and possible causes; however, it has a high missed alarm rate, lacks the ability to identify unknown attacks, and requires the management of a massive signature database. Anomaly detection is built on the idea of developing a normal behaviour model and then identifying abnormal behaviours based on how much they deviate from that profile. As a consequence, when applying an anomaly detection system, it is important to define a consistent profile. Anomaly detection has strong generalizability and the ability to diagnose unexplained attacks, but it has a high false alarm rate and fails to have possible reasons for anomalies. The key differences between misuse detection and anomaly detection as seen in Table 1.

Table 1. Differences between misuse detection and anomaly detection.

	Misuse Detection	Anomaly Detection
Detection performance	Low false alarm rate; High missed alarm rate	Low missed alarm rate; High false alarm rate
Detection efficiency	High, decrease with scale of signature database	Dependent on model complexity
Dependence on domain knowledge	Almost all detections depend on domain knowledge	Low, only the feature design depends on domain knowledge
Interpretation	Design based on domain knowledge, strong interpretative ability	Outputs only detection results, weak interpretative ability
Unknown attack detection	Only detects known attacks	Detects known and unknown attacks

As seen in Figure 1, misuse recognition approaches include pattern matching-based, expert system-based, and finite state machine-based techniques. Anomalies was discovered using

statistical model-based, machine learning-based, and time series-based methods..

CONCLUSION

This research paper includes the details of a real-world testbed that we created to conduct cyber-attacks and design an intrusion detection system (IDS). We study computer attacks and demonstrate how a machine learning-based anomaly detection technology can detect them effectively. We used representative tests to quantify the methods' effectiveness in order to provide a practical estimate of their efficacy.

REFERENCES

1. This research paper includes the details of a real-world testbed that we created to conduct cyber-attacks and design an intrusion detection system (IDS). We study computer attacks and demonstrate how a machine learning-based anomaly detection technology can detect them effectively. We used representative tests to quantify the methods' effectiveness in order to provide a practical estimate of their efficacy.
2. Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Tamilnadu, India, 2019, pp. 1-3, doi: 10.1109/INCOS45849.2019.8951344.
3. M. S. Koli and M. K. Chavan, "An advanced method for detection of botnet traffic using intrusion detection system," 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 2017, pp. 481-485, doi: 10.1109/ICICCT.2017.7975246.
4. D. Subramanyam, "Classification of Intrusion Detection Dataset using machine learning Approaches," 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Belgaum, India, 2018, pp. 280-283, doi: 10.1109/CTEMS.2018.8769270.
5. R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat and S. Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," in IEEE Access, vol. 7, pp. 41525-41550, 2019, doi: 10.1109/ACCESS.2019.2895334.
6. M. Raihan-Al-Masud and H. A. Mustafa, "Network Intrusion Detection System Learning," 2019 IEEE International Conference on Telecommunications and Photonics (ICTP), Dhaka, Bangladesh, 2019, pp. 1-4, doi: 10.1109/ICTP48844.2019.9041736.
7. G. Karatas, O. Demir and O. K. Sahingoz, "Increasing the Performance of Machine Learning-Based IDSs on an Imbalanced and Up-to-Date Dataset," in IEEE Access, vol. 8, pp. 32150-32162, 2020, doi: 10.1109/ACCESS.2020.2973219.
8. B. W. Masduki, K. Ramli, F. A. Saputra and D. Sugiarto, "Study on implementation of machine learning methods combination for improving attacks detection accuracy on Intrusion Detection System (IDS)," 2015 International Conference on Quality in Research (QiR), Lombok, Indonesia, 2015, pp. 56-64, doi: 10.1109/QiR.2015.7374895.
9. D. H. Lakshminarayana, J. Philips and N. Tabrizi, "A Survey of Intrusion Detection Techniques," 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, 2019, pp. 1122-1129, doi: 10.1109/ICMLA.2019.00187.
10. L. Deng and D. Gao, "Research on Immune Based Adaptive Intrusion Detection System Model," 2009 International Conference on Networks Security, Wireless Communications and Trusted Computing, Wuhan, China, 2009, Using Voting Ensemble Machine pp. 488-

491, doi: 10.1109/NSWCTC.2009.87.

11. M. A. Jabbar, R. Aluvalu and S. S. Satyanarayana Reddy, "Intrusion Detection System Using Bayesian Network and Feature Subset Selection," 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, India, 2017, pp. 1-5, doi: 10.1109/ICCIC.2017.8524381.
12. M. Azizjon, A. Jumabek and W. Kim, "1D CNN based network intrusion detection with normalization on imbalanced data," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, 2020, pp. 218-224, doi: 10.1109/ICAIIC48513.2020.9064976.
13. T. Poongothai and K. Duraiswamy, "Intrusion detection in mobile AdHoc networks using machine learning approach," International Conference on Information Communication and Embedded Systems (ICICES2014), Chennai, India, 2014, pp. 1-5, doi: 10.1109/ICICES.2014.7033949.
14. W. Wang, X. Du, D. Shan, R. Qin and N. Wang, "Cloud Intrusion Detection Method Based on Stacked Contractive Auto-Encoder and Support Vector Machine," in IEEE Transactions on Cloud Computing, doi: 10.1109/TCC.2020.3001017.