Review Article

ACCENTS Transactions on Information Security, Vol 5(19)
ISSN (Online): 2455-7196
http://dx.doi.org/10.19101/TIS.2020.517002

An analytical survey on the role of machine learning algorithms in case of intrusion detection

Anand Vijay[1*], Kailash Patidar[2], Manoj Yadav[2] and Rishi Kushwah[2]
M.Tech Scholar, Department of Computer Science, School of Engineering, Sri Satya Sai University of Technology & Medical Sciences, Sehore, Madhya Pradesh, India[1]
Assistant Professor, Department of Computer Science, School of Engineering, Sri Satya Sai University of Technology & Medical Sciences, Sehore, Madhya Pradesh, India[2]

Abstract
In this paper an analytical survey on the role of machine learning algorithms in case of intrusion detection has been presented and discussed. This paper shows the analytical aspects in the development of efficient intrusion detection system (IDS). The related study for the development of this system has been presented in terms of computational methods. The discussed methods are data mining, artificial intelligence and machine learning. It has been discussed along with the attack parameters and attack types. This paper also elaborates the impact of different attack and handling mechanism based on the previous papers.

Keywords
Intrusion detection system, Attack types, Data mining, Artificial intelligence, Machine learning.

1.Introduction
Intrusion detection is an important aspect where there is the need of computational techniques like data mining, artificial intelligence and machine learning [1−4] for the improvement in detection system. It has been found that different algorithms have already been applied in the same direction for the improvement [3−8]. But there are several areas where there is the need of improvement including detection, identification along with the attack types.

Based on the literature it has been found that the intrusion detection is an important aspect in different areas of data sharing and communication [9]. It may be helpful in the identification of malicious and suspicious behavior. It has been done through intrusion detection system (IDS) [10]. These systems have been developed to identify suspicious activities which may be attack prone or it may increase the chances of vulnerable activities [11−14].

Other important aspects are types of intrusion, types of attacks, identification process and the detection process. Detection process includes network, configuration, IP, signature, host, anomaly and configuration.

Figure 1 shows the intrusion detection system process.

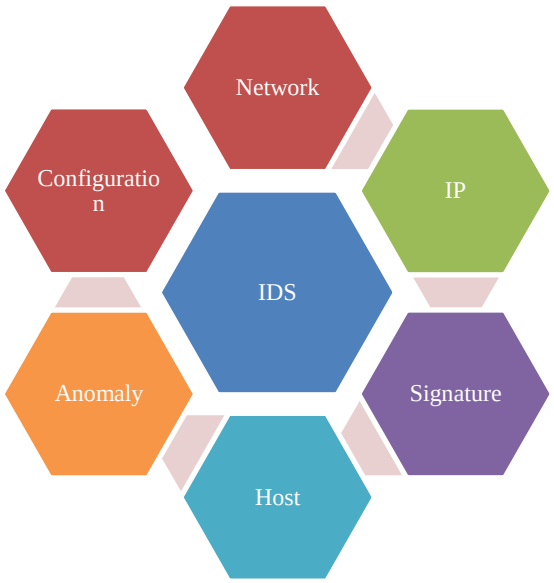

Figure 1 Intrusion detection system process

The main objective of this paper are as follows:
1. To explore the computational methodologies for the efficient intrusion detection system development.

*Author for correspondence

2. To explore the parametric evaluation of data mining, artificial intelligence and machine learning algorithms for the intrusion detection process.
3. To explore different attacks and the possibilities in the cyber security.
4. To explore the host, configuration and anomaly parameters in terms of detection framework for different intrusions.

2.Literature survey

In 2020, Razimi et al. [15] discussed about the surveillance technology. They have proposed an intelligent home surveillance system. IT has been proposed based on the use of Raspberry Pi. It has been triggered when an intruder is captured through the video surveillances.

In 2020, Zoppi et al. [16] discussed about the anomaly detection techniques. It has been discussed in terms of identifying patterns. Their main aim is to instruct the anomaly-based techniques considering unsupervised algorithms. It has been used for the classification of normal and anomalous behaviors.

In 2020, Dang [17] discussed about intrusion detection system. The main task of the detection system is to differentiate benign and malicious network flows. They have discussed the active learning usage. It has been discussed in terms of active learning for the online configuration. It has been discussed for the reduction of labeling cost.

In 2020, Chen et al. [18] discussed about the 5G application and the chances of intrusion detection. They have suggested that the traditional method is relatively insufficient. They have proposed a RLA intrusion detection system for the hybrid network. For the classification support vector machine algorithm has been used. They have achieved 98% accuracy approximately.

In 2020, Jin et al. [19] discussed about the applicability of big data and machine learning algorithms in case of intrusion detection. They have proposed a K-nearest neighbors (KNN) and categorical boosting (CatBoost) for the imbalanced data. For experimentation they have used KDD99 dataset. By this method they have achieved better detection performance.

In 2019, Halimaa and Sundarakantham [20] dicusses about malicious activity and intrusion detection system. They have suggested that the intrusion detection may plays an important role in the network. Hey have suggested the need of classification methodologies. They have applied support vector machine (SVM) and naïve Bayes (NB) algorithm for the classification problem. For experimentation NSL-KDD dataset has been used. Their result suggest that the support vector machine outperforms.

In 2020, Taghavinejad et al. [21] discussed the use of Internet of Things. They have discussed regarding the prevention from the cyber-attack through intrusion detection system. They have used the combination of SVM, KNN and decision tree (DT). Their result shows that the proposed method is found to be better.

In 2020, Mu et al. [22] discussed about the internet intrusion detection. They have discussed the applicability in terms of IP matching and network monitoring. They have also discussed unauthorized access due to various tags.

In 2020, Dawit et al. [23] discussed about cyber security. They have investigated several methods for the intrusion detection collaboration. They have also studied the integration of intrusion detection. They have also studied and discussed the major vulnerabilities in case of blockchain application.

In 2020, Park et al. [24] discussed a prediction model which is based on recurrent neural network. They have discussed this in terms of IoT environment. They have used long short-term memory model. They have used cosine similarity for the scoring function. They have considered a normal packet for the same.

In 2020, Iman and Ahmad [25] discussed about the intrusion detection system development. They have analyses and estimated the use of random forest algorithm. They have considered Boruta algorithm. Their results show that the proposed method is capable of preventing the infinite loop. It is capable in the improvement of the performance.

3.Discussion and comparative analysis

Based on the previous literature it has been found that different approaches have been used for the intrusion detection system. Based on the literature it has also been found that there is the need of classification algorithms. *Table 1* shows regarding different approaches and its applicability.

Table 1 Different approaches and its applicability

S. No	Reference	Approach Used	Applicability	Results
1	[26]	XGBoost Classification	They have proposed a hybrid principal component analysis (PCA)-firefly based machine learning model. It has been used for the intrusion detection system datasets classification. They have collected the data for Kaggle.	Their result shows that their proposed model performs better than machine learning models.
2	[27]	Hybrid deep learning model	They have proposed a hybrid deep learning model. It has been proposed for the efficient detection. It is based on convolutional neural network.	Their result shows improved performance in terms of traditional methods.
3	[28]	Deep learning approaches	They have survey and analyzes deep learning approaches.	They have included future research challenges and future directions.
4	[29]	Deep learning	They have analyzed different deep learning models. They have also studied the performance of binary and multiclass classification. They have considered traffic datasets, namely, the CSE-CIC-IDS2018 and the Bot-IoT dataset for the experimentation.	They have evaluated the performances based on accuracy, false alarm rate.
5	[30]	Feature selection algorithm	They have proposed wrapper feature selection algorithm for the IDS. They have applied pigeon inspired optimizer. It has been applied for the selection process. They have considered KDDCUP99, NLS-KDD and UNSW-NB15 datasets.	The proposed algorithm has been evaluated based on accuracy, F-score, etc. Their result also shows faster convergence than the sigmoid method.

4.Problem identification

After the current trend's discussion and analysis, the following gaps have been identified:

1. There is a need of categorization and clustering techniques to handle large amount of data.
2. Better intrusion detection framework can be created based on data mining and evolutionary techniques.
3. Most of the algorithms performs well on some attacks like DoS but fails in prediction of different attacks.
4. Classification accuracy can be improved separately for each attack along with the average accuracy.
5. Most of the algorithms calculated the accuracy on the selected data so there is the need of classifying complete data.

5.Conclusion

In this paper a survey and analysis has been presented based on the computational methods for efficient IDS development. Different method has been considered for the review and analysis. Mostly data mining, machine learning and artificial intelligence algorithms have been considered for the analysis. As per the security concern it is very important to secure the connection and timely detection of intrusion. The aim of this paper is to provide a better insight in the direction of intrusion detection and find the implications of different methodology as far presented.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] McLaughlin S, Konstantinou C, Wang X, Davi L, Sadeghi AR, Maniatakos M, Karri R. The cybersecurity landscape in industrial control systems. Proceedings of the IEEE. 2016; 104(5):1039-57.

[2] Ani UP, He H, Tiwari A. Review of cybersecurity issues in industrial critical infrastructure: manufacturing in perspective. Journal of Cyber Security Technology. 2017; 1(1):32-74.

[3] Gupta R, Singh S. A review on intrusion detection system based on data mining and evolutionary algorithms. International Journal of Advanced Technology and Engineering Exploration. 2018; 5(46):356-61.

[4] Kim S, Kim B, Kim HJ. Intrusion detection and mitigation system using blockchain analysis for bitcoin exchange. In proceedings of the 2018 international conference on cloud computing and internet of things 2018 (pp. 40-4).

[5] Ren W, Yardley T, Nahrstedt K. EDMAND: Edge-based multi-level anomaly detection for SCADA networks. In international conference on communications, control, and computing technologies

for smart grids (SmartGridComm) 2018 (pp. 1-7). IEEE.

[6] Kumar KN, Sukumaran S. A survey on network intrusion detection system techniques. International Journal of Advanced Technology and Engineering Exploration. 2018; 5(47):385-93.

[7] Yang J, Shen C, Chi Y, Xu P, Sun W. An extensible Hadoop framework for monitoring performance metrics and events of OpenStack cloud. In 3rd international conference on big data analysis (ICBDA) 2018 (pp. 222-6). IEEE.

[8] Foroushani ZA, Li Y. Intrusion detection system by using hybrid algorithm of data mining technique. In proceedings of the 2018 7th international conference on software and computer applications 2018 (pp. 119-23).

[9] Farhaoui Y. How to secure web servers by the intrusion prevention system (IPS)?. International Journal of Advanced Computer Research. 2016; 6(23):65.

[10] Sicard F, Zamaï É, Flaus JM. An approach based on behavioral models and critical states distance notion for improving cybersecurity of industrial control systems. Reliability Engineering & System Safety. 2019; 188:584-603.

[11] Anwer HM, Farouk M, Abdel-Hamid A. A framework for efficient network anomaly intrusion detection with features selection. In international conference on information and communication systems (ICICS) 2018 (pp. 157-62). IEEE.

[12] Alexopoulos N, Vasilomanolakis E, Ivánkó NR, Mühlhäuser M. Towards blockchain-based collaborative intrusion detection systems. In international conference on critical information infrastructures security 2017 (pp. 107-18). Springer, Cham.

[13] Kaushik M, Ojha G. Attack penetration system for SQL injection. International Journal of Advanced Computer Research. 2014; 4(2):724-32.

[14] Lopez-Martin M, Carro B, Sanchez-Esguevillas A. Application of deep reinforcement learning to intrusion detection for supervised problems. Expert Systems with Applications. 2020; 141:112963.

[15] Razimi UN, Alkawaz MH, Segar SD. Indoor intrusion detection and filtering system using raspberry Pi. In international colloquium on signal processing & Its applications (CSPA) 2020 (pp. 18-22). IEEE.

[16] Zoppi T, Ceccarelli A, Bondavalli A. Into the unknown: unsupervised machine learning algorithms for anomaly-based intrusion detection. In annual IEEE-IFIP international conference on dependable systems and networks-supplemental volume (DSN-S) 2020 (pp. 81-81). IEEE.

[17] Dang QV. Active learning for intrusion detection systems. In Research, Innovation and Vision for the Future 2020.

[18] Chen W, Cao H, Lv X, Cao Y. A hybrid feature extraction network for intrusion detection based on global attention mechanism. In international conference on computer information and big data applications (CIBDA) 2020 (pp. 481-5). IEEE.

[19] Jin D, Lu Y, Qin J, Cheng Z, Mao Z. KC-IDS: multi-layer intrusion detection system. In international conference on high performance big data and intelligent systems (HPBD&IS) 2020 (pp. 1-5). IEEE.

[20] Halimaa A, Sundarakantham K. machine learning based intrusion detection system. In international conference on trends in electronics and informatics (ICOEI) 2019 (pp. 916-20). IEEE.

[21] Taghavinejad SM, Taghavinejad M, Shahmiri L, Zavvar M, Zavvar MH. Intrusion detection in IoT-based smart grid using hybrid decision tree. In international conference on web research (ICWR) 2020 (pp. 152-6). IEEE.

[22] Mu Z, Liu H, Liu C. Design and implementation of network intrusion detection system. In international conference on intelligent transportation, big data & smart city (ICITBS) 2020 (pp. 494-7). IEEE.

[23] Dawit NA, Mathew SS, Hayawi K. Suitability of blockchain for collaborative intrusion detection systems. In annual undergraduate research conference on applied computing (URC) 2020 (pp. 1-6). IEEE.

[24] Park SH, Park HJ, Choi YJ. RNN-based prediction for network intrusion detection. In international conference on artificial intelligence in information and communication (ICAIIC) 2020 (pp. 572-74). IEEE.

[25] Iman AN, Ahmad T. Improving intrusion detection system by estimating parameters of random forest in boruta. In international conference on smart technology and applications (ICoSTA) 2020 (pp. 1-6). IEEE.

[26] Bhattacharya S, Kaluri R, Singh S, Alazab M, Tariq U. A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU. Electronics. 2020.

[27] Hassan MM, Gumaei A, Alsanad A, Alrubaian M, Fortino G. A hybrid deep learning model for efficient intrusion detection in big data environment. Information Sciences. 2020; 513:386-96.

[28] Aldweesh A, Derhab A, Emam AZ. Deep learning approaches for anomaly-based intrusion detection systems: a survey, taxonomy, and open issues. Knowledge-Based Systems. 2020; 189:105124.

[29] Ferrag MA, Maglaras L, Moschoyiannis S, Janicke H. Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study. Journal of Information Security and Applications. 2020; 50:102419.

[30] Alazzam H, Sharieh A, Sabri KE. A feature selection algorithm for intrusion detection system based on pigeon inspired optimizer. Expert Systems with Applications. 2020; 148:113249.