ISSN 2347 - 3983

Volume 8. No. 4, April 2020

International Journal of Emerging Trends in Engineering Research

Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter01842020.pdf

https://doi.org/10.30534/ijeter/2020/01842020

A Survey on Intrusion Detection Mechanism using Machine Learning Algorithms

Pavithra G[1], Abirami P[2], Bhuvaneshwari S[3], Dharani S[4], Haridharani B[5]

[1]Department of CSE, M.Kumarasamy College of Engineering, Karur, India, pavithrag.cse@mkce.ac.in
[2]Department of CSE, M.Kumarasamy College of Engineering, Karur, India, abiramipandiankrr@gmail.com
[3]Department of CSE, M.Kumarasamy College of Engineering, Karur, India, bhuvanakumarsiva1995@gmail.com
[4]Department of CSE, M.Kumarasamy College of Engineering, Karur, India, dharanis1811@gmail.com
[5]Department of CSE, M.Kumarasamy College of Engineering, Karur, India, macintosh.haridharani@gmail.com

ABSTRACT

The use of internet is getting rapidly increased which lead to various security issues in the network. Attackers continuously develop new approach to track the valuable information from the users. IDS has been evolved to detect suspicious attack effectively. Various techniques are used for finding intrusion. In this paper, the survey is presented on intrusion detection system. The survey was about the techniques that have been used in intrusion detection system and complete knowledge about the strengths and limitations of detection methods which provides a foundation for developing efficient intrusion detection system.

Key words: Internet, security, intrusion detection, network, attacks

1. INTRODUCTION

The wide spread of internet all over the world has lead to various security issues in the network. The security of computer system is at risk due to maximum usage of internet. The rapid growth of network related issues was analyzed frequently. Many researchers have been done to detect attacks. Network security is important for Protecting the user's data and resources. IDS (Intrusion detection system) are used for finding any unwanted activity on the system. IDS can be a hardware or software appliance used to monitor traffic on the internet and used to identify threats. Various methods have been introduced for controlling unauthorized access to system. The system proposed to detect this intrusion is called IDS (Intrusion Detection System). This system is used to decide whether the activities done by the unauthorized user is intrusive or normal based on the available resources. The administrators are used to detect the violations in IDS. The methods commonly used are anomaly and misuse detection in IDS. Anomaly detection used for finding new attacks and misuse detection find known attacks. Various threats are rapidly developing and valuable information of the users can be stolen by the hackers. As various attacks have been detected in the network, a system needs to be designed to prevent these attacks. Hence IDS is evolved to satisfy that requirement. The accuracy in detection rate is the main issue and so many machine learning (ML) algorithms are used to detect attacks. The basic element of IDS is data object. The information related to the attacks are carried by the data object. Alert should be provided when any malicious activity is detected. So that by giving the alert, authorized users can able to know the threats. Some attacks are very complex to handle and this system provide the way for effective IDS (Intrusion Detection System).

2. LITERATURE SURVEY

Fangfei Weng [1] found a new intrusion detection system [IDS] to improve the detection rate using clustering ensemble. It is used to avoid mistaken clustering, so that detection rate efficiency was improved. They use KDD-99 dataset to display the efficiency. The data collected is based on the probability value. This paper focus mainly on data collection stage. The dataset is found in this paper was accurate and this is important to find whether the dataset is valuable or not. The attacks get grouped under four categories and one drawback is finding threshold value. In future, implementation can be done for detecting various types of attacks.

Meijuan Gao [2] uses the concept of Support Vector Machine [SVM] for improving the convergence value and accuracy is also flexible. By overcoming the disadvantages of BP network, SVM are proposed.

This method classifies the data by using the hyper plane in optimal way. In IDS system, stages include data collection, preprocessing and collection stage. The dataset get split as testing data and training data. This paper has full focus on classification stage. While using this method, it provides more accuracy and low false positive rate. Different training samples are used every time to get accurate result. Better results are provided by support vector machine and further work can be done by hybrid methods and good accuracy can be achieved.

Pingjie Tang [3] aimed to design the intrusion detection system using different types of algorithms for improving the accuracy effectively. In process step, various symbols and non numeric data are available and it should get converted. The attacks are categorized under five steps. Triangle area is calculated in this paper. The edges in triangle are calculated by distance formula. This system is based on combination of clustering and classification for getting good performance. In later work, better feature selection algorithm has to get provided for detection.

LI Han [4] proposed a concept to find the attacks in the computer network. The rate of detection is increased in this paper. MDKM (Modified Dynamic K-means) algorithm provides the way to find unwanted activities like noise. The distance among the samples is calculated. KDD (Knowledge Discovery Dataset) CUP 1999 data set provides high accuracy and low false alarm rate [FAR] and also used to detect four attacks such as Neptune, buffer overflow, guess password, port sweep. In this paper, dynamic K means algorithm is discussed but here also, finding K value is difficult. The future work is to improve the accuracy for large data set and repeated data should get rectified.

Roshan Chitrakar [5] discovered an approach to increase the accuracy and detect the attacks in the network. This approach combines K- medoids and SVM algorithm for improving the accuracy and used to maintain standard clustering in K-medoids. The same instance is grouped under K- medoids method. The data get separated and distance is calculated and low cost is applied. The separation of data is done by using classification stage and taken attributes get separated into training and testing values. ROC techniques are practiced for accuracy process. This approach uses the Kyoto2006+ data set for enhancing the accuracy. Further work on this system helps to improve accuracy when using kernel based SVM and also reduce the time complexity of K-medoids.

Yi Gong [6] proposed multi agent intrusion detection system. This paper used to overcome the network security issue by using feature selection approach. Prevention control can be done by using large switched network. Detecting reliability is getting increased in this paper. Feature selection is efficient because it removes irrelevant noisy features. This let to better results for intrusion detection. Monitor, communication and decision agents are used in this paper. The results of this paper show that BN classifier works more efficiently. The improvement in industrial control system can be done in future work.

Mohsen Eslamnezhad [7] have introduced a new approach in intrusion detection system to improve detection rates in various types of attacks and minimize the false positive rate [FPR] using Min-Max method. Hence, the algorithm is compared to K-means clustering algorithm. The output of this paper provides higher rate of detection for normal, probe and denial of service. Whereas, U2R and R2L resulting a lower detection rate. Here, they are using advanced KDD data set to enhance the performance ratio and rate of finding detection is increased. The algorithm used in this paper finds the unknown attacks also. Therefore, the method has been more efficient for intrusion detection system. The further work of this paper is to improve the efficiency rate for unknown or new attacks.

Piyush A.Sonewar [8] focuses on web attacks that are used by hackers to steal information. The author describes and gives importance to crucial attacks such as SQL injection and XSS attacks. The injection attack causes changes in the database which can affect the backend of the application. For SQL checking, query is checked for patterns that have been matched. If the matching is found, the query will pass or the attack is detected. The alerts have been given based on an outcome of the model. The future work can be done to reduce web threats in efficient manner.

Varuna [9] have discussed about intrusion detection system to safeguard entire networking system. For this, different methods have been used. Normally, we have used Misuse detection. The already used methods have several failures, so they were introducing a new technique called Hybrid Classifiers, which gives an improved accuracy rate and find other detection attacks in IDS. While introducing hybrid method, it combines both (K-means clustering algorithm) and NB classifier (Naive Bayes) which improves new features and finds distance relation between data set to a number of centroids. While using other algorithms, it results a greater performance rate for only Dos attacks and normal. In this paper, the efficient results are producing on Probe, U2R, R2L attacks by using naïve bayes algorithm.

Saba Khan [10] discusses the topic on threats that are found in web. Both signature and anomaly based IDS is proposed. Application Layer is focused and the attacks that are happening in application layer is rectified. The unwanted queries can be injected through sql attack and malicious script can be attached by Cross-site attacks. Genetic approach is used for the detection of attacks by converting input data into binary code. Log data is created for attacks that are found in the application. The author focus on some crucial attacks and future work can be done on IDS to detect web attacks effectively.

Victor Clincy [11] proposed work on injection attacks on web services. Genetic algorithm is used for finding web defects. The new set of signatures is found with existing one. The defects that are discussed here are injection attack, that retrieve the data from the database table and X path is the attack in which data is retrieved from xml document. Request is sent to service provided and that in return gives response and repository is made for log. The xml messages are provided between web servers and firewall that is examined with signature. The future work can be done on anomaly based IDS.

Anand Sukumar J V [12] introduced a new approach to determine the attacks in computer network. The accuracy of the system is improved by IGKM method and they compare K-means ++ and IGKM to prove that IGKM provides more accuracy (72.91%) than the K-means algorithm. The dataset is clustered and k value is provided by fitness function. For large datasets, time complexity is low. The dataset used in the system is KDD-99 dataset. So, there may be a chance for repeated data sets. Four types of attacks are discussed in this paper. The disadvantage of the approach is that when we are using small data set, it gives less accuracy. This can be solved in future.

Anish Halimaa [13] has researched to solve the malicious activity. For that, we are introducing a new method is Intrusion Detection System. The method called SVM (Support Vector Machine) and Naive Bayes algorithm. In this paper they are comparing two algorithms for finding an accuracy rate and misclassification rate. NSL-KDD dataset provides high efficiency. The major issue in this technique is to resolve huge traffic in network and security system. The results shows high for Support Vector Machine [SVM] compared to Naive Bayes. The Support Vector Machine shows accuracy rate for 19000 instances. Whereas, Naive Bayes shows higher misclassification rate than SVM. The future work results that multi-level hybrid model to be used for large volume of data set for high accuracy.

Table 1: Comparison of various Machine Learning Algorithms for Intrusion Detection Mechanism

S.No	Year	Author of the paper	Algorithm	Results (%)
1	2007	Fangfei Weng [1]	K-means	FPR - 0.7736
2	2009	Meijuan Gao [2]	Support Vector Machine	MSE for training sample - 0.0085 MSE for testing sample - 0.0213
3	2010	Pingjie Tang [3]	Triangle Area Support Vector Machine	FAR - 2.99 Precision % Probe - 95.43 U2R - 86.21 R2L - 93.96

4	2011	LI Han [4]	Modified Dynamic K-means	FAR DoS - 1.7 U2R - 2.8 R2L - 0.4 Probe - 1.9
5	2012	Roshan Chitrakar [5]	Support Vector Machine K-medoids	FPR – 0.46
6	2014	Yi Gong [6]	MinMax K – means	FPR- 9
7	2014	Mohsen Eslamnezhad [7]	Feature selection approach	FPR PROBE - 0.3 U2R - 0.1 R2L - 0.1
8	2015	Piyush A.Sonewar [8]	Mapping Model	Alert has been given for SQL and XSS
9	2015	Varuna [9]	K-means Naive Bayes	DR Normal - 74.11 DoS - 86.05 Probe - 92.48 R2L- 32.02 U2R- 19.0
10	2017	Saba Khan [10]	Evolutionary Algorithm	Finding SQL and XSS attacks for both anomaly and signature based IDS
11	2017	Victor Clincy [11]	Genetic Algorithm	Selection rate-20 Mutation rate- 0.3
12	2018	Anand Sukumar [12]	Improved Genetic K-means	Accuracy- 72.91 TP- 6 FP- 17 FN- 6
13	2019	Anish Halimaa [13]	Support Vector Machine Naive Bayes	Misclassification rate-2.705

3. CONCLUSION

The survey of this paper shows the evolution of research regarding Intrusion detection system and various methods and machine learning algorithms that has been proposed for finding distinct attacks. However the current system for detecting intrusion has several drawbacks. So the future work of this survey needs to build a system that would recognize unknown attacks effectively with better accuracy.

4. REFERENCES

1. F. Weng, Q. Jiang, L. Shi and N. Wu, "**An Intrusion Detection System Based on the Clustering Ensemble**," 2007 International Workshop on Anti-Counterfeiting, Security and Identification (ASID), Xiamen, Fujian, 2007, pp. 121-124.

2. M. Gao, J. Tian and M. Xia, "**Intrusion Detection Method Based on Classify Support Vector**

Machine," 2009 Second International Conference on Intelligent Computation Technology and Automation, Changsha, Hunan, 2009, pp. 391-394.
https://doi.org/10.1109/ICICTA.2009.330

3. P. Tang, R. Jiang and M. Zhao, "**Feature Selection and Design of Intrusion Detection System Based on k-Means and Triangle Area Support Vector Machine**," 2010 Second International Conference on Future Networks, Sanya, Hainan, 2010, pp. 144-148.

4. L. Han, "**Using a Dynamic K-means Algorithm to Detect Anomaly Activities**," 2011 Seventh International Conference on Computational Intelligence and Security, Hainan, 2011, pp. 1049-1052.
https://doi.org/10.1109/CIS.2011.233

5. R. Chitrakar and H. Chuanhe, "**Anomaly detection using Support Vector Machine classification with k-Medoids clustering**," 2012 Third Asian Himalayas International Conference on Internet, Kathmandu, 2012, pp. 1-5.

6. Y. Gong, Y. Fang, L. Liu and J. Li, "**Multi-agent Intrusion Detection System Using Feature Selection Approach**," 2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kitakyushu, 2014, pp. 528-531.

7. M. Eslamnezhad and A. Y. Varjani, "**Intrusion detection based on MinMax K-means clustering**," 7'th International Symposium on Telecommunications (IST'2014), Tehran, 2014, pp. 804-808.
https://doi.org/10.1109/ISTEL.2014.7000814

8. P. A. Sonewar and N. A. Mhetre, "**A novel approach for detection of SQL injection and cross site scripting attacks**," 2015 International Conference on Pervasive Computing (ICPC), Pune, 2015, pp. 1-4.

9. S. Varuna and P. Natesan, "**An integration of k-means clustering and naïve bayes classifier for Intrusion Detection**," 2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN), Chennai, 2015, pp. 1-5.

10.S. Khan and D. Motwani, "**Implementation of IDS for web application attack using evolutionary algorithm**," 2017 International Conference on Intelligent Computing and Control (I2C2), Coimbatore, 2017, pp. 1-5.

11. V. Clincy and H. Shahriar, "**Web service injection attack detection**," 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST), Cambridge, 2017, pp. 173-178.

12. J. V. Anand Sukumar, I. Pranav, M. Neetish and J. Narayanan, "**Network Intrusion Detection Using Improved Genetic k-means Algorithm**," International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 2441-2446.
https://doi.org/10.1109/ICACCI.2018.8554710

13. A. Halimaa A. and K. Sundarakantham, "**Machine Learning Based Intrusion Detection System**," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 916-920.

14. Praveen Kumar Kollu and R.Satya Prasad "**Intrusion Detection System Using Recurrent Neural Networks and Attention Mechanism**" International Journal of Emerging Trends in Engineering Research, Volume 7, No. 8 August 2019, ISSN 2347 – 3983.
https://doi.org/10.30534/ijeter/2019/12782019

15. N.Chandra Sekhar Reddy, Purna Chandra Rao Vemuri and A.Govardhan "**An Emperical Study on Support Vector Machines for Intrusion Detection**" International Journal of Emerging Trends in Engineering Research (IJETER) ,Vol.7,No.10,World Academy of Research in Science and Engineering (WARSE)2019.
https://doi.org/10.30534/ijeter/2019/037102019

16. E.T. Venkatesh, P. Thangaraj, and S.Chitra " **An Improved Neural Approach for Malignant and Normal Colon Tissue Classification from Oligonucleotide Arrays**" European Journal of Scientific Research , vol. 54 , pp. 159 – 164 , 2011.

17. S.Thilagamani, N. Shanthi "**Object Recognition Based on Image Segmentation and Clustering**" Journal of Computer Science,Vol.7,No.11,pp. 1741-1748, 2011.

18.P.Santhi, G.Mahalakshmi, **Classification of Magnetic Resonance Images Using Eight Directions Gray Level Co-Occurrence Matrix Based Feature Extraction**, International Journal of Engineering and Advanced Technology, ISSN: 2249-8958, Volume-8 Issue-4, April 2019.

19.P. Pandiaraja, N Deepa 2019 , **A Novel Data Privacy-Preserving Protocol for Multi-data Users by using genetic algorithm** , Journal of Soft Computing , Springer , Volume 23 ,Issue 18, Pages 8539-8553.

20.K Sumathi, P Pandiaraja , **Dynamic alternate buffer switching and congestion control in wireless multimedia sensor networks** , Journal of Peer-to-Peer Networking and Applications , Springer.

21.P.RajeshKanna and P.Pandiaraja 2019, **An Efficient Sentiment Analysis Approach for Product Review using Turney Algorithm** , Journal of Procedia Computer Science , Elsevier ,Vol 165 , Issue 2019, Pages 356-362.
https://doi.org/10.1016/j.procs.2020.01.038