International Journal of Advanced Technology and Engineering Exploration, Vol 3(25)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2016.325005

Review Article

A survey an analysis for an efficient intrusion detection system

Neeru Jain[1*] and Deepak Tomar[2]
M.Tech Scholar, Department of Computer Science and Engineering, TIT, Bhopal[1]
Professor, Department of Computer Science and Engineering, TIT, Bhopal[2]

Abstract
The process of identifying possible attacks in the network is called intrusion detection. As per the security concern it is very important to secure the connection and timely detection so that any fraud / unauthorized event will not be happened. 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. Different data mining techniques along with the evolutionary algorithms have been analyzed so that a better framework can be designed based on the hybridization of efficient algorithms. For this purpose denial of service (DoS), user to root (U2R), remote to user (R2L) and probe attacks have been considered in this paper.

Keywords
Intrusion detection, Data mining, Evolutionary algorithms, DoS, U2R, R2L and Probe.

1.Introduction
There are several research work is in progress in the intrusion detection which is based on data mining techniques [1]. Several security techniques and recommendation are suggested in different literatures [2-6]. It is the process of identifying possible attacks in the network. Intrusion identification is the procedure of malevolent attacks from the network and system when it is currently correspondence or removing information in the constant environment [2, 3]. Since its creation, interruption identification has been one of the key components in accomplishing data security. It goes about as the second-line protection which supplements the get to controls. At the point when the controls fizzled, the identification frameworks ought to have the capacity to recognize it constant and caution the security officers to take quick and suitable activities [3-8].

Interruption identification framework manage directing the episodes happening in PC framework or system situations and looking at them for indications of conceivable occasions, which are encroachment or inescapable dangers to PC security, or standard security hones Intrusion recognition frameworks (IDS) have developed to identify activities which jeopardize the uprightness, secrecy or accessibility of are source as a push to give an answer for existing security issues [9].

There are several problems which can be identified like data preprocessing in the huge network or huge node list. So how to handle the dataset is also an important task [10]. Several security schemes are suggested concurrently to securing the data in different literatures [11-15]. The main aim of this paper is to find a solution or a hybrid framework based on data mining and evolutionary computation to improve the efficiency of intrusion identification. These methods are useful and have been used in different papers with different attack detection strategy [16-20].

2.Related work
In 2012, Prasenna et al. [21] proposed that in routine system security basically depends on numerical calculations and low counter measures to taken to counteract interruption discovery framework, albeit the vast majority of this methodologies as far as hypothetically tested to actualize. Creators recommend that as opposed to producing expansive number of tenets the advancement streamlining methods like genetic network programming (GNP) can be utilized .The GNP depends on coordinated diagram. They concentrate on the security issues identified with send an information mining-based IDS in a constant domain. They sum up the issue of GNP with affiliation run mining and propose a fluffy weighted affiliation run mining with GNP system reasonable for both nonstop and discrete qualities.

*Author for correspondence

In 2011, Han [22] concentrates on interruption identification in light of grouping examination. The point is to enhance the recognition rate and reduction the false alert rate. A modified dynamic K-means algorithm called MDKM to identify inconsistency exercises is proposed and relating reenactment tests are introduced. Firstly, the MDKM calculation channels the commotion and detached focuses on the information set. Besides by computing the separations between all example information focuses, they get the high-thickness parameters and bunch parcel parameters, utilizing dynamic iterative process we get the k grouping focus precisely, then an abnormality identification model is introduced. They utilized KDD CUP 1999 information set to test the execution of the model. Their outcomes demonstrate the framework has a higher location rate and a lower false alert rate, it accomplishes hopeful point.

In 2014, Benaicha et al. [23] exhibit a genetic algorithm (GA) approach with an enhanced introductory populace and determination administrator, to effectively recognize different sorts of system interruptions. They utilized GA to streamline the inquiry of assault situations in review records, on account of its great adjust investigation/misuse; as indicated by the creators it gives the subset of potential assaults which are available in the review document in a sensible preparing time. The testing is on NSL-KDD99 benchmark dataset has been utilized to distinguish the abuse exercises. Their approach of IDS with Genetic calculation builds the execution of the recognition rate of the Network Intrusion Detection Model and diminishes the false positive rate.

In 2014, Thaseen et al. [24] proposed a novel strategy for principle component analysis (PCA) and support vector machine (SVM) by streamlining the piece parameters utilizing programmed parameter choice method. Their approach diminishes the preparation and testing time to distinguish interruptions in this manner enhancing the precision. Their proposed technique was tried on KDD information set. The datasets were precisely partitioned into preparing and testing considering the minority assaults, for example, U2R and R2L to be available in the testing set to distinguish the event of obscure assault. Their outcomes show that the proposed technique is fruitful in recognizing interruptions.

In 2014, Wagh et al. [25] recommended Network security is a vital part of web empowered frameworks in the present world situation. As per the creators because of mind boggling chain of PCs the open doors for interruptions and assaults have expanded. In this way it is need of great importance to locate the most ideal routes conceivable to secure our frameworks. So the creators propose interruption identification framework is assuming crucial part for PC security. The best technique used to tackle issue of IDS is machine learning. Thy watched that the rising field of semi administered learning offers a guaranteed path for corresponding exploration. So they proposed a semi-administered technique to lessen false caution rate and to enhance location rate for IDS.

In 2014, Sayar et al. [26] suggested that the associating with web can be both worthwhile and disadvantageous it might be said that web can give to such an extent solace to business furthermore colossal hazard to end clients. Increment in the speed of data information stream furthermore advancement in correspondence organize alongside numerous variables there is plausibility of number of assaults on PC framework. Keeping in mind the end goal to shield PC framework from these attacks and vindictive exercises interruption recognition framework came into picture. They give us a frame of intrusion detection and discussed different methods used to execute interruption location framework.

In 2011, Islam et al. [27] suggested that the intrusion recognition framework in remote sensor system is one of the developing examination territories as of late. Intrusion discovery is one of the critical angles for remote sensor systems. There are two distinctive sort of interruption location system: oddity based and signature based as suggested by the authors. They specify a few attacks on WSN and principally concentrate just on the oddity based interruption location framework. At long last, we talk about around a few existing ways to deal with depict how they have distinguished security dangers and actualized their interruption discovery framework.

In 2015, Bahl et al. [28] suggested U2R attack class is an open research problem. Their purpose of this study is to identify the important features to improve the detection rate and reduce the false detection rate. The researched highlight subset choice methods enhance the general exactness, discovery rate of U2R attack class furthermore decrease the computational expense. The experimental results have demonstrated a discernible change in location rate of U2R attack class with highlight subset determination systems. In

2015, Yan et al. [29] proposed an intelligent intrusion detection model. Taking into account the attributes of worldwide predominance of hereditary calculation and area of nerve, the model upgrades the weights of the neural system utilizing genetic algorithm. Test results demonstrate that the insightful way can enhance the proficiency of the interruption identification.

In 2015, Haidar et al. [30] emphasizes the significance of abnormality based interruption recognition procedures, the vital results of these frameworks, most recent created techniques and what is normal from the future trials in this field. In addition, the method of learning client profiles impacts in recognizing interruptions can be explored. Finally, the lights will be shed on an offline approach using Multi-Layer Perceptron (MLP) and Self Organizing Maps (SOM) which is a distinguished method in intrusion detection.

3.Gap statements

After the current trends 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.

4.Comparative analysis

Table 1 shows the results comparison of the traditional approaches for checking the classification accuracy.

Table 1 Results comparison for different methods

S.NO	Method used	Classification accuracy obtained (%)	Attacks included
1	Intrusion detection using genetic algorithm [23]	99.74	DoS
		NA	U2R
		NA	R2L
		NA	Probe
2	Intrusion detection model using fusion of PCA and optimized SVM [24]	99.78	DoS
		NA	U2R
		NA	R2L
		NA	Probe
3	Intrusion detection based on k-means clustering and Naïve Bayes classification [31]	99.5	DoS
		40	U2R
		61.6	R2L
		100	Probe
4	Naïve Bayes classifier [32]	99.9	DoS
		NA	U2R
		NA	R2L
		NA	Probe
5	Artificial Immune System Approach [33]	96.79	DoS
		NA	U2R
		NA	R2L
		NA	Probe
6	Random Forest Modeling [34]	99.67	DoS
		99.67	U2R
		99.67	R2L
		99.67	Probe
7	Supervised Machine Learning Algorithms [35]	99.0	DoS
		NA	U2R
		NA	R2L
		NA	Probe
8	chi-square feature selection and multi class SVM [36]	99.9	DoS
		73.9	U2R
		98.7	R2L

S.NO	Method used	Classification accuracy obtained (%)	Attacks included
		99.2	Probe

5.Conclusions and recommendations

In this paper several methods for intrusion detection have been analyzed and discussed. The main focus of this paper is to discuss the detection mechanism in two ways first to elaborate the accuracy efficiency obtained by the different approaches and second the type of attacks covered by the methodology. Based on the discussion and analysis the following recommendations have been suggested:

1. Need of including all the attacks and checking the overall accuracy obtained in all the attacks.
2. Hybridization of different techniques can be useful as there is a need of classification, categorization and clustering.
3. To handle large amount of data in a single slot is also a big and challenging problem.
4. Computation time is also important in classification.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] Jianliang M, Haikun S, Ling B. The application on intrusion detection based on k-means cluster algorithm. In international forum on information technology and applications 2009 (pp. 150-2). IEEE.
[2] Sharma N, Gaur B. An approach for efficient intrusion detection for KDD dataset: a survey. International Journal of Advanced Technology and Engineering Exploration. 2016; 3(18): 72-6.
[3] Lundin E, Jonsson E. Survey of intrusion detection research. Chalmers University of Technology; 2002.
[4] Tian L, Jianwen W. Research on network intrusion detection system based on improved k-means clustering algorithm. In international forum on computer science-technology and applications 2009 (pp. 76-9). IEEE.
[5] Conteh NY, Schmick PJ. Cybersecurity: risks, vulnerabilities and countermeasures to prevent social engineering attacks. International Journal of Advanced Computer Research. 2016; 6(23):31-38.
[6] Farhaoui Y. How to secure web servers by the intrusion prevention system (IPS)? International Journal of Advanced Computer Research. 2016; 6(23):65-71.
[7] Ishida M, Takakura H, Okabe Y. High-performance intrusion detection using optigrid clustering and grid-based labelling. In international symposium on applications and the internet 2011 (pp. 11-9). IEEE.
[8] Devaraju S, Ramakrishnan S. Performance analysis of intrusion detection system using various neural network classifiers. In international conference on recent trends in information technology 2011 (pp. 1033-8). IEEE.
[9] Brugger ST. Data mining methods for network intrusion detection. University of California at Davis. 2004.
[10] Sirisha GN, Shashi M. Subspace clustering for high dimensional datasets. International Journal of Advanced Computer Research. 2016; 6(26):177-184.
[11] Murugavalli S, Jainulabudeen SA, Kumar GS, Anuradha D. Enhancing security against hard AI problems in user authentication using CAPTCHA as graphical passwords. International Journal of Advanced Computer Research. 2016; 6(24):93-9.
[12] Lee W, Stolfo SJ. Data mining approaches for intrusion detection. In Usenix security 1998.
[13] Nalavade K, Meshram BB. Mining association rules to evade network intrusion in network audit data. International Journal of Advanced Computer Research. 2014; 4(2):560-7.
[14] Naoum R, Aziz S, Alabsi F. An enhancement of the replacement steady state genetic algorithm for intrusion detection. International Journal of Advanced Computer Research. 2014; 4(2):487-93.
[15] Lee W, Stolfo SJ, Mok KW. A data mining framework for building intrusion detection models. In proceedings of the IEEE symposium on security and privacy 1999 (pp. 120-32). IEEE.
[16] Tiwari R, Sinhal A. Block based text data partition with RC4 encryption for text data security. International Journal of Advanced Computer Research. 2016; 6(24):107-13.
[17] Sperotto A, Schaffrath G, Sadre R, Morariu C, Pras A, Stiller B. An overview of IP flow-based intrusion detection. IEEE Communications Surveys & Tutorials. 2010; 12(3):343-56.
[18] Li Z, Li Y, Xu L. Anomaly intrusion detection method based on k-means clustering algorithm with particle swarm optimization. In international conference on information technology, computer engineering and management sciences 2011 (pp. 157-61). IEEE.
[19] Manimaran A, Durairaj M. The conjectural framework for detecting DDoS attack using enhanced entropy based threshold technique (EEB-TT) in cloud environment. International Journal of Advanced Computer Research. 2016; 6(27):230.
[20] Yin-huan LI. Design of intrusion detection model based on data mining technology. In 2012 international conference on industrial control and electronics engineering 2012.
[21] Prasenna P, Kumar RK, Ramana AR, Devanbu A. Network programming and mining classifier for

intrusion detection using probability classification. In international conference on pattern recognition, informatics and medical engineering 2012 (pp. 204-9). IEEE.

[22] Han LI. Using a dynamic k-means algorithm to detect anomaly activities. In seventh international conference on computational intelligence and security (CIS) 2011 (pp. 1049-52). IEEE.

[23] Benaicha SE, Saoudi L, Guermeche SE, Lounis O. Intrusion detection system using genetic algorithm. In science and information conference (SAI), 2014 (pp. 564-8). IEEE.

[24] Thaseen IS, Kumar CA. Intrusion detection model using fusion of PCA and optimized SVM. In international conference on contemporary computing and informatics 2014 (pp. 879-84). IEEE.

[25] Wagh SK, Kolhe SR. Effective intrusion detection system using semi-supervised learning. In international conference on data mining and intelligent computing 2014 (pp. 1-5). IEEE.

[26] Sayar AA, Pawar SN, Mane V. A review of intrusion detection system in computer network. International Journal of Computer Science and Mobile Computing. 2014; 3(2):700-3.

[27] Islam MS, Rahman SA. Anomaly intrusion detection system in wireless sensor networks: security threats and existing approaches. International Journal of Advanced Science and Technology. 2011; 36(1):1-8.

[28] Bahl S, Sharma SK. Improving classification accuracy of intrusion detection system using feature subset selection. In fifth international conference on advanced computing & communication technologies 2015 (pp. 431-6). IEEE.

[29] Yan C. Intelligent intrusion detection based on soft computing. In seventh international conference on measuring technology and mechatronics automation 2015 (pp. 577-80). IEEE.

[30] Haidar GA, Boustany C. High perception intrusion detection system using neural networks. In international conference on complex, intelligent, and software intensive systems 2015 (pp. 497-501). IEEE.

[31] Muda Z, Yassin W, Sulaiman MN, Udzir NI. Intrusion detection based on k-means clustering and naïve Bayes classification. In international conference on information technology in Asia 2011 (pp. 1-6). IEEE.

[32] Deshmukh DH, Ghorpade T, Padiya P. Intrusion detection system by improved preprocessing methods and naïve Bayes classifier using NSL-KDD 99 Dataset. In international conference on electronics and communication systems 2014 (pp. 1-7). IEEE.

[33] Khalkhali I, Azmi R, Azimpour-Kivi M, Khansari M. Host-based web anomaly intrusion detection system, an artificial immune system approach. International Journal of Computer Science. 2011; 8(5): 14-24.

[34] Farnaaz N, Jabbar MA. Random forest modeling for network intrusion detection system. Procedia Computer Science. 2016; 89:213-7.

[35] Belavagi MC, Muniyal B. Performance evaluation of supervised machine learning algorithms for intrusion detection. Procedia Computer Science. 2016; 89:117-23.

[36] Thaseen IS, Kumar CA. Intrusion detection model using fusion of chi-square feature selection and multi class SVM. Journal of King Saud University-Computer and Information Sciences. 2016.