Review Article

International Journal of Advanced Technology and Engineering Exploration, Vol 5(46)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2018.545024

A review on intrusion detection system based on data mining and evolutionary algorithms

Ravindra Gupta[1*] and Shailendra Singh[2]
PhD Research Scholar, Department of Computer Science and Engineering, BU, Bhopal, Madhya Pradesh, India[1]
Professor and Senior Member IEEE, Department of Computer Engineering and Application, N.I.T.T.T.R., Bhopal, Madhya Pradesh, India[2]

Abstract

Intrusion detection is the procedure for determining intrusions in the network. This paper explores the methodology in the direction of intrusion detection system. It explores the possibility of enhancement and propounding the advantages. This study helps in exploring the method analytically, methodically and experimentally. This paper lists the gaps and the advantages, so that future framework can be design to enhance the efficiency. It also provides the detail discussion based on the attributes and parameters variations. Finally future suggestions have been listed.

Keywords

***Data mining, Evolutionary algorithms, Intrusion detection, Network system*.**

1.Introduction

Different methodological research has been already done and in progress [1]. Different methods have been used and applied. In this paper we have mainly focused on data mining and evolutionary algorithm based detection. It is the technique of pernicious assaults from the system and framework when it is as of now correspondence or expelling data in the steady condition [2, 3]. Since its creation, interference recognizable proof has been one of the key parts in achieving information security. It goes about as the second-line insurance which supplements the get to controls. Exactly when the controls failed, the ID systems should have the ability to remember it steady and alert the security officers to take speedy and reasonable exercises [3−8].

Interference recognizable proof structure oversee coordinating the scenes occurring in PC system or framework circumstances and taking a gander at them for signs of possible events, which are infringement or certain threats to PC security, or standard security sharpens intrusion detection systems (IDS) have created to distinguish exercises which risk the uprightness, mystery or availability of are source as a push to give a response for existing security issues [9].

There are a few issues which can be recognized like information pre-processing in the enormous system or gigantic hub list. So how to deal with the dataset is likewise an essential assignment [10]. A few security plans are recommended simultaneously to anchoring the information in various written works [11−15].

The fundamental point of this paper is to discover a hybrid system in light of information mining and transformative calculation to enhance the effectiveness of interruption distinguishing proof [16−20]. These techniques are valuable and have been utilized in various papers with various attack recognition system [21, 22].

2.Related work

In 2014, Benaicha et al. [23] presented genetic algorithm (GA) along with the amendments initial population and selection. It is used in the optimization of the search of attack scenarios in audit files. It gives the subset of potential assaults which are available in the review document in a sensible preparing time. They have used the network security laboratory-knowledge discovery and data mining (NSL-KDD99) dataset. By combining the IDS with Genetic algorithm increases the performance of the detection rate of the network intrusion detection model and reduces the false positive rate.

[*] Author for correspondence

In 2014, Thaseen and Kumar [24] suggested that in old IDS methods there is the problem of high false alarm rate. They have suggested that the use of machine learning algorithms may improve the performance. In this work a principal component analysis (PCA) and support vector machine (SVM) based hybrid method. It is used for the optimization of the kernel parameters using automatic parameter selection technique. This procedure diminishes the preparation and testing time to distinguish interruptions consequently enhancing the exactness. Their technique was tried on KDD informational collection. The datasets were deliberately separated into preparing what's more, trying considering the minority assaults, for example, U2R and R2L to be available in the testing set to recognize the event of obscure assault. The outcomes show that the work strategy is effective in recognizing interruptions. The exploratory outcomes demonstrate that the arrangement exactness of the work technique outflanks other grouping procedures utilizing SVM as the classifier and other dimensionality diminishment or highlight determination systems. Least assets are devoured as the classifier input requires lessened list of capabilities and subsequently limiting preparing and testing overhead time.

In 2014, Wagh and Kolhe [25] suggested that there is need to protect the systems efficiently. They have suggested that there are several attacks have been notice regularly. The most powerful strategy used to take care of issue of IDS is machine learning. Getting marked information does not just require additional time be that as it may, it is additionally costly. Marked information alongside unlabeled information is utilized in semi-administered strategies. The rising field of semi supervised learning offers a guaranteed path for corresponding inquires about. In this paper, a successful semi-regulated technique to diminish false alert rate and to enhance recognition rate for IDS.

In 2014, Sayar et al. [26] recommended the development of web world is approaching an individual yet at same time there is a danger of being robed. Associating with web can be both beneficial and disadvantageous one might say that web can give as much solace to business and furthermore huge hazard to end clients. Increment in the speed of data information stream and furthermore advancement in correspondence organizes 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 assaults and vindictive exercises interruption location framework came into picture. They also provide us diagram of interruption recognition framework and different strategies used to actualize interruption identification framework.

In 2015, Bahl and Sharma [27] suggested that the IDS have grown rapidly. They have suggested that the user to root (U2R) attack detection is open research in the IDS system. Current IDS utilizes all information highlights to distinguish interruptions. A portion of the highlights might be excess to the identification procedure. The motivation behind this experimental examination is to recognize the essential highlights to enhance the discovery rate and diminish the false discovery rate. The explored highlight subset determination methods enhance the general precision, discovery rate of U2R assault class and furthermore decrease the computational cost. The exact outcomes have demonstrated a discernible change in recognition rate of U2R assault class with highlight subset choice methods.

In 2015, Yan [28] shows an intelligent intrusion detection model. In view of the attributes of worldwide prevalence of hereditary calculation and territory of nerve, the model streamlines the weights of the neural system utilizing hereditary calculation. Investigation results demonstrate that the astute way can enhance the proficiency of the interruption identification.

In 2015, Haidar and Boustany [29] suggested that the anomaly-based network intrusion detection is important against malicious acts. They have focused irregularity based intrusion recognition strategies, the critical results of these frameworks, most recent created strategies and what is normal from what's to come tests in this field. In addition, the procedure of learning client profiles impacts in distinguishing interruptions will be talked about. At last, the lights will be shed on a disconnected approach utilizing multi-layer perceptron (MLP) and self-organizing maps (SOM).

In 2017, Kumar et al. [30] shows an improved fuzzy membership function to detect anomalies and intrusions. The goal of the present approach is to accomplish an ideal change grid which can enhance classifier correctness’s. The change lattice is gone for mapping the first process onto another fluffy space; with the goal that the resultant portrayal is free from commotion information and encourages moving forward the general precision and furthermore singular class correctness’s. Trial results demonstrate that correctness's acquired utilizing our approach is better contrasted with different methodologies. Specifically U2R and R2L correctness's are recorded to be in particular promising. This examination demonstrates an approach which addresses the change in generally speaking exactness and furthermore change in recognizing R2L and U2R assault correctness's.

In 2017, Ding and Wang [31] suggested the need of IDS which is able to prevent attacks. Profound learning has been turned out to be the most productive strategy to identify the intrusions. They have shown a deep neural network (DNN) model to identify the anomalies. The model is fundamentally made out of multi-layer completely associated layer and dropout layer. Adam calculation is utilized in the model to anticipate the identification show from falling into neighborhood least and speed up preparing speed. Rectified linear unit (ReLU) has been used as the activation function in each layer as the input layer and softmax is used as the output layer. It is applied on the KDD CUP 99 dataset. Reenactment results demonstrate that the execution of the model is superior to alternate models.

In 2017, Xiaofeng and Xiaohong [32] suggested an efficient approach based on k-means and multi-level SVM. K-means algorithm is used to cluster based on data detected. Then multi-level SVM to stamp the unusual group for itemized order, the last acknowledgment of the recognition of system assaults. This work interruption recognition calculation utilizes the NSL-KDD informational index to mimic the analysis. The outcomes demonstrate that the calculation can enhance the system interruption identification rate and decrease the false caution rate. It is a successful method for organize security assurance.

In 2017, Potteti and Parati [33] described the hybrid IDS which is based on fuzzy genetic algorithm. They have suggested the main drawback of IDS is high rate of false positive. By planning a crossover interruption location framework can tackle this by associating a location module to the irregularity discovery module. Their hybrid intrusion detection system for wireless local area networks. It is based on fuzzy genetic logic. The fuzzy genetic logic-based system could be capable to recognize the nosy exercises of the PC systems as the control base holds a superior arrangement of tenets.

In 2017, Balasaraswathi et al. [34] suggested that the IDS routinely handles monstrous measures of information movement that contain repetitive and superfluous highlights, which affect the execution of the IDS adversely. Highlight choice strategies assume a critical part in taking out inconsequential and repetitive highlights in IDS. Factual examination, neural systems, machine learning, information mining strategies, and bolster vector machine models are utilized in some such techniques. They have suggested that the better classification accuracy can be achieved through feature selection. They have surveyed in this direction.

In 2017, Shah et al. [35] suggested that the intrusion detection system is a classifier which gathers confirmations for the nearness of interruption and raises an alert for any variations from the norm exhibit. Be that as it may, the utilization of interruption discovery framework experiences two noteworthy disadvantages: higher false alert rate and lower location rate; these breaking point the recognition execution of interruption identification framework. An imminent approach for enhancing execution is using numerous sensors/interruption location frameworks. Confirmation hypothesis is a numerical hypothesis of proof which is utilized to intertwine confirmations from various wellsprings of confirmation and yields a worldwide choice. The work in this paper examines the impediments and issues with confirm hypothesis and proposes an altered structure for combination of alerts of various intrusion detection frameworks.

In 2018, Almi'ani et al. [36] suggested that the impact of information security breaching is the crucial aspects now days. New and more refined assaults are rising and created; requiring the data frameworks and systems be ensured in an exceedingly adaptable and precise way. They have used artificial neural networks for addressing the high accuracy and precision demands. They have built an intelligent IDS based on clustered version of SOM network. The framework comprises of two resulting stages: first, SOM arrange was fabricated, at that point a various leveled agglomerative bunching utilizing k-implies was connected on SOM neurons. The work in this examination paper tends to the issues of affectability and time utilization for every association record handling. This framework was shown utilizing NSL-KDD benchmark dataset, where it has accomplished better affectability came to up than 96.66 % in under 0.08 milliseconds for each association record.

In 2018, Anwer et al. [37] suggested that the machine learning algorithms for the detection in anomalies using supervised and unsupervised approaches. A framework for efficient network anomaly interruption detection with features selection. They have presented a features selection framework for anomaly detection by the help of machine learning classifiers. The system applies distinctive procedures by utilizing channel and wrapper highlights choice approaches. The point of this structure is to choose the base number of highlights that accomplish the most astounding exactness. UNSW-NB15 dataset is utilized in the trial results to assess the structure.

3.Comparative analysis

The comparative analysis based on the previous results has been shown in *Table 1*. This analysis provides us the detail about the method used approach applied and the attack considered. It will be helpful in finding the current insights in the intrusion detection area.

Table 1 Results comparison for different methods

S. No.	Source	Method Used	Approach	Attack detected
1	Pamukov and Poulkov [38]	Multiple negative selection algorithm	Their algorithm is based on negative selection algorithm and the co-stimulation principles. They have used two-tiered negative selection process. Their aim is to decrease the number of detection errors. *They have suggested that there is the need of variability in the data size and the identification of the optimized sets. The classification accuracy obtained in their approach is 90%.*	Negative detectors
2	Desai and Gaikwad [39]	Real time hybrid intrusion detection system	They have implemented hybrid intrusion detection system. It is capable in identifying the internal and external both type of attacks. For internal attacks signature matching have been used. For external attack detection fuzzy genetic algorithm have been applied. *They have suggested the need of more than one algorithm for the efficient intrusion detection. The classification accuracy obtained in their approach is 93%.*	Overall detection
3	Maske and Parvat [40]	Advanced anomaly intrusion detection technique	This paper introduces a method for deducing the call traces for the raw system. They have suggested that the improved results may obtained by making use of range of decision engines. *They have suggested the need of decision based ranking for the efficient intrusion detection classification. The classification accuracy obtained in their approach is 88%.*	Overall detection
4	Garg and Maheshwari [41]	Snort-based intrusion	Snort is open source software and used mostly used in signature based IDS. It is utilized world broadly in intrusion identification and counteractive action space. *They have suggested different approach with different dataset. The classification accuracy obtained in their approach is 91%.*	Overall detection
5	Mehmood and Rais [42]	Machine learning algorithms for intrusion detection	They have suggested that the signature based method may fail in the detection of the novel attacks. They have compared supervised algorithms. *They have suggested the need of randomize selection and hybridization of different approaches with different lable. The classification accuracy obtained in their approach is 96%.*	DoS, Probe, R2L and U2R
6	Gupta et al. [43]	Data mining techniques for the network intrusion detection	They have applied different data mining algorithms like linear regression and k-means to classify network activities. *They have suggested the need of feature selection to select the most relevant features which can be helpful in the detection. The classification accuracy obtained in their approach is 81%.*	Overall detection

After the current trends discussion and analysis the following gaps have been identified:

1) Different mechanism can be hybridized for better attack detection accuracy.

2) The combination of data mining and optimization algorithms can be used.
3) Need to concentrate on the individual attack type and their sub-type.
4) It is found that the accuracies for each attack like DoS, U2R, R2L and Probe should be discussed along with the combination.
5) There is the need of level wise clustering and supervising the approach with the optimization set.

4.Conclusion

In this study several previous methods have been analyzed and discussed. Based on the study of different approaches have been explored and analyzed. Different classification and clustering algorithms have been analyzed. Based on this the following analysis, future suggestions have been listed:

Individual classification accuracy based on the attacks can be analyzed. This will provide a way to design and develop a framework for individual attack classification mechanism.

Combination of data mining and optimization techniques may be useful. It will be helpful in analyzing, categorizing along with the matching optimization for the better accuracy. Execution time handling in the larger dataset is also a big challenge.

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] Mohamed MH, Waguih HM. A proposed academic advisor model based on data mining classification techniques. International Journal of Advanced Computer Research. 2018; 8(36):129-36.

[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):31-8.

[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] Irandegani M, Bagherizadeh M. Designing an asynchronous multi-channel media access control protocol based on service quality for wireless sensor networks. International Journal of Advanced Computer Research. 2017; 7(32):190-9.

[8] Devaraju S, Ramakrishnan S. Performance analysis of intrusion detection system using various neural network classifiers. International conference on recent trends in information technology (ICRTIT) 2011 (pp. 1033-8).

[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):177-84.

[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 symposium 1998 (pp. 79-93).

[13] Nalavade K, Meshram BB. Mining association rules to evade network intrusion in network audit data. International Journal of Advanced Computer Research. 2014; 4(15):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(15):487-94.

[15] Lee W, Stolfo SJ, Mok KW. A data mining framework for building intrusion detection models. In proceedings of the 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 and 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):230-7.

[20] Yin-huan LI. Design of intrusion detection model based on data mining technology. In international conference on industrial control and electronics engineering 2012 (pp. 571-4). IEEE.

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

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

[29] 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.

[30] Kumar GR, Mangathayaru N, Narsimha G, Reddy GS. Evolutionary approach for intrusion detection. In international conference on engineering & MIS 2017 (pp. 1-6). IEEE.

[31] Ding S, Wang G. Research on intrusion detection technology based on deep learning. In international conference on computer and communications 2017 (pp. 1474-8). IEEE.

[32] Xiaofeng Z, Xiaohong H. Research on intrusion detection based on improved combination of K-means and multi-level SVM. In international conference on communication technology 2017 (pp. 2042-5). IEEE.

[33] Potteti S, Parati N. Intrusion detection system using hybrid fuzzy genetic algorithm. In international conference on trends in electronics and informatics 2017 (pp. 613-8). IEEE.

[34] Balasaraswathi VR, Sugumaran M, Hamid Y. Feature selection techniques for intrusion detection using non-bio-inspired and bio-inspired optimization algorithms. Journal of Communications and Information Networks. 2017; 2(4):107-19.

[35] Shah V, Aggarwal AK, Chaubey N. Performance improvement of intrusion detection with fusion of multiple sensors. Complex & Intelligent Systems. 2017; 3(1):33-9.

[36] Almi'ani M, Ghazleh AA, Al-Rahayfeh A, Razaque A. Intelligent intrusion detection system using clustered self-organized map. In international conference on software defined systems 2018 (pp. 138-44). IEEE.

[37] 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 2018 (pp. 157-62). IEEE.

[38] Pamukov ME, Poulkov VK. Multiple negative selection algorithm: improving detection error rates in IoT intrusion detection systems. In international conference on intelligent data acquisition and advanced computing systems: technology and applications 2017 (pp. 543-7). IEEE.

[39] Desai AS, Gaikwad DP. Real time hybrid intrusion detection system using signature matching algorithm and fuzzy-GA. In international conference on advances in electronics, communication and computer technology 2016 (pp. 291-4). IEEE.

[40] Maske SA, Parvat TJ. Advanced anomaly intrusion detection technique for host based system using system call patterns. In international conference on inventive computation technologies 2016 (pp. 1-4). IEEE.

[41] Garg A, Maheshwari P. Performance analysis of snort-based intrusion detection system. In international conference on advanced computing and communication systems 2016 (pp. 1-5). IEEE.

[42] Mehmood T, Rais HB. Machine learning algorithms in context of intrusion detection. In international conference on computer and information sciences 2016 (pp. 369-73). IEEE.

[43] Gupta D, Singhal S, Malik S, Singh A. Network intrusion detection system using various data mining techniques. In international conference on research advances in integrated navigation systems 2016 (pp. 1-6). IEEE.

Ravindra Gupta is a PhD Research Scholar at BU, Bhopal Madhya Pradesh in the department of Computer Science and Engineering and has a Master degree in Computer Science and Engineering from RGPV, Bhopal, Madhya Pradesh. His area of interests are Soft Computing, Data Mining and Network Security.
Email: ravindra_p84@rediffmail.com