An efficient intrusion detection system based on data mining and evolutionary algorithm

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

Received: July 04, 2018 Accepted: August 20, 2018

ABSTRACT *This paper proposed an efficient approach for intrusion detection which is based on data mining and evolutionary algorithm. The standard NSL-KDD dataset has been used for the experimentation and comparative study. This includes the data separation along with the k-clusters. So that proper classification has been applied and the classification accuracy can be improved. The parameters used are content, traffic and the host. Particle swarm optimization has been applied based on the maximum threshold and the objective function. The results shows that the proposed approach has 97% average accuracy in case of denial of service (DoS), user to root (U2R), remote to user (R2L) and probe attacks.*

Keywords: *NSL-KDD, PSO, DoS, U2R, R2L, Probe*

1. Introduction

There are a few research work is in advancement in the interruption location which depends on information mining systems [1]. A few security strategies and proposal are recommended in various literary works [2−6]. It is the way toward distinguishing conceivable assaults in the system.

Interruption ID is the strategy of noxious assaults from the system and framework when it is at present correspondence or evacuating data in the steady condition [2, 3]. Since its creation, intrusion distinguishing proof has been one of the key parts in achieving information security. It goes about as the second-line assurance which supplements the get to controls. Exactly when the controls failed, the distinguishing proof systems should have the ability to remember it consistent and alert the security officers to take speedy and reasonable exercises [3−8]. Interruption distinguishing proof structure oversee coordinating the scenes occurring in PC structure or framework circumstances and taking a gander at them for signs of possible events, which are infringement or certain risks to PC security, or standard security sharpens Intrusion identification systems (IDS) have created to recognize exercises which imperil the uprightness, mystery or openness of are source as a push to give a response for existing security issues [9].

There are a few issues which can be distinguished like information preprocessing in the gigantic system or tremendous hub list. So how to deal with the dataset is likewise an imperative errand [10]. A few security plans are recommended simultaneously to anchoring the information in various written works [11−15].

The primary point of this paper is to discover an answer or a half and half system dependent on information mining and developmental calculation to enhance the effectiveness of interruption ID. These techniques are valuable and have been utilized in various papers with various assault discovery procedures [16−22].

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 auditfiles. 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 theperformance of the detection rate of the network intrusiondetection model and reduces the false positive rate.

In 2014, Thaseen et al. [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 supportvector machine (SVM) based hybrid method. It is used for the optimization of the kernel parameters usingautomatic parameter selection technique.

This procedure diminishes the preparation and testing time to distinguish interruptions consequentlyenhancing 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 et al. [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 supervisedlearning 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.

In 2015, Bahl et al. [27] suggested that the IDS havegrown 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 et al. [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 utilizingmulti-layer perceptron (MLP) and self-organizing maps(SOM).

In 2017, Kumar et al. [30] shows an improvedfuzzy 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 et al. [31] suggested the needof IDSwhich 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 theactivation function in each layer as the input layer and softmax is used as theoutput 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 et al. [32] suggested an efficient approach based on k-means andmulti-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 et al. [33] described the hybrid IDS which is based on fuzzy genetic algorithm. They have suggested the main drawback of IDS is high rate of falsepositive. 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 wirelesslocal area networks. It is based on fuzzy genetic logic. Thefuzzy 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 framework. 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 precisiondemands. They have built anintelligent IDS based on clustered version ofSOM 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 usingsupervised 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 machinelearning 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. Method

This approach is developed on the NETBEANS IDE environment supported with the JDK version 7 or higher. It supports the data either to select it randomly or the whole data simultaneously. For the experimentation random data has been considered as for the comparison purpose to supporting comparison from the previous research work. Although there is an option for selecting all the data simultaneously.

At that point we consider the ordinary information set and for discovering the interruptions we ascertain coordinating variable. In the first step the data NSL-KDD cup99 is divided into clusters based on the filtration parameters that are content feature, traffic features and the host feature. In the event that the worth crosses the farthest point esteem then the hub will be included into the last unsafe class. Then PSO have been applied on the cluster for the further data classification. The whole iteration is random so there is no biasness. At long last taking into account the DoS, U2R, R2) and Probing (Probe) class to locate the final classification. This can be better understood from the flowchart shown in figure 1.

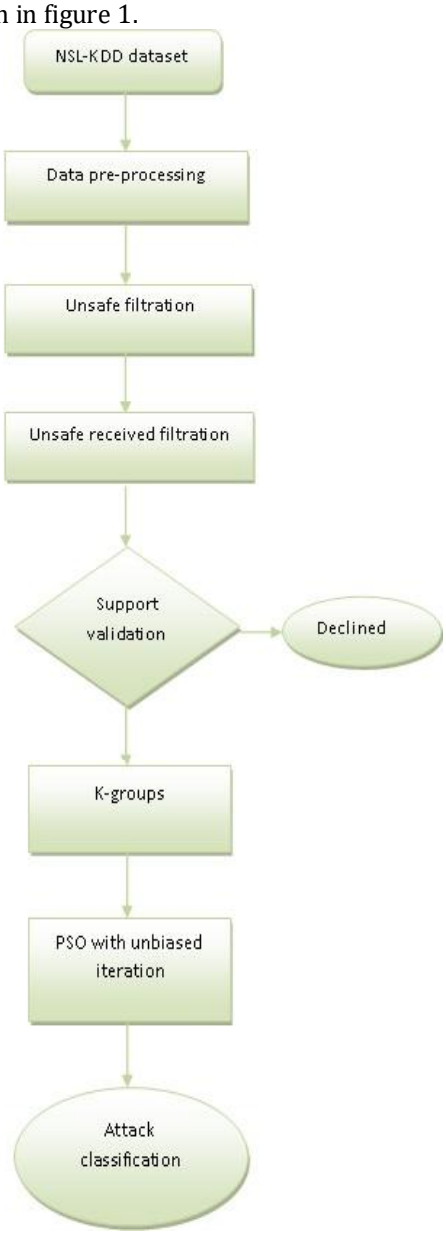

Figure 1: Flowchart of my Proposed Technique

The classified data is sending the output to the input of PSO.

PSO Algorithm

Input: PSOIS (psois1, psois2, psois3…….psosn)

Output: OS (os1, os2, os3……osn)

Terminology:

PSOIS: PSO input system

PSOISu: PSO input system updated

Ns: Next span

OS: Output system

Kw: K produced weight

NS: Next span

Rv: Random value

Rvp: Previous random value

Step 1: Data is loaded from the K-output set

Step 2: Consider loaded data as psois1, psois2, psois3…….psosn for the final data classification.

Step 3:

3a. Five span data are considered.

3b. for i=1 to 5 do

$Kw=\sum psois1^* Rv + psois2^* Rv + psois3^* Rv$ ….psois5* Rv /n

for 2 to 5 do

$Ns2=Kw+\sum psois1u * Rv + psois2u * Rv + psois3u * Rv$ ….psois5u * Rv /n - Rvp

$Ns3= Ns2+\sum acois1u * Rv + acois2u * Rv + acois3u * Rv$ ….acois5u * Rv /n- Rvp

$Ns4= Ns3+\sum acois1u * Rv + acois2u * Rv + acois3u * Rv$ ….acois5u * Rv /n- Rvp

$Ns5= Ns4+\sum acois1u * Rv + acois2u * Rv + acois3u * Rv$ ….acois5u * Rv /n- Rvp

If(Nsn+1>Nsn)

Nsn+1 = Nsn

else

No change.

Step 4: Repeat the section 3 till all the iterations are not completed.

Step 5: Final classified outputs have been achieved.

4. Result

In this section the results obtained from our method has been discussed. Figure 1-4 shows the results of our approach. For the result discussion different ranges have been considered. The average classification accuracy is shown in figure 2. The results shows that the average classification accuracy is better in comparison to others and it is improved in the classification of all the attacks.

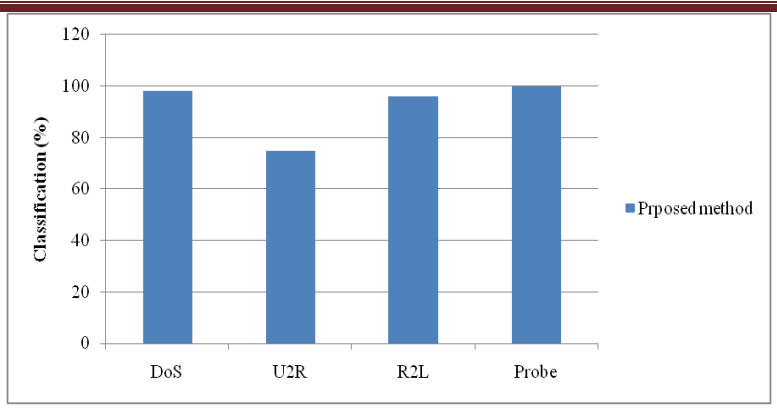

Figure 1: Accuracy comparison in case of Random selection-1classification comparison

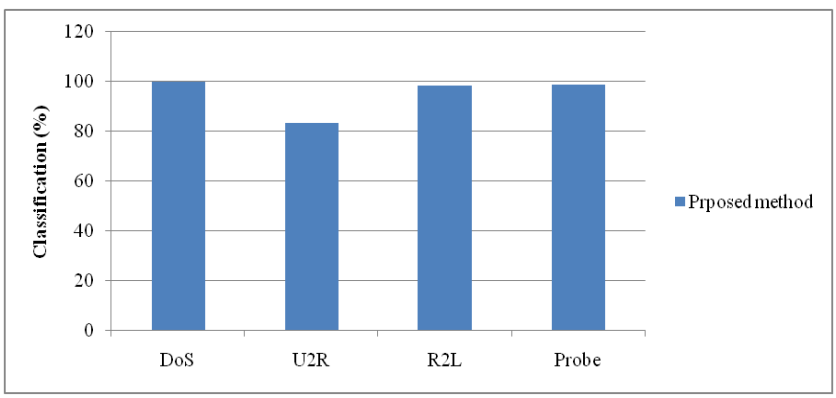

Figure2: Accuracy comparison in case of Random selection-2 classification comparison

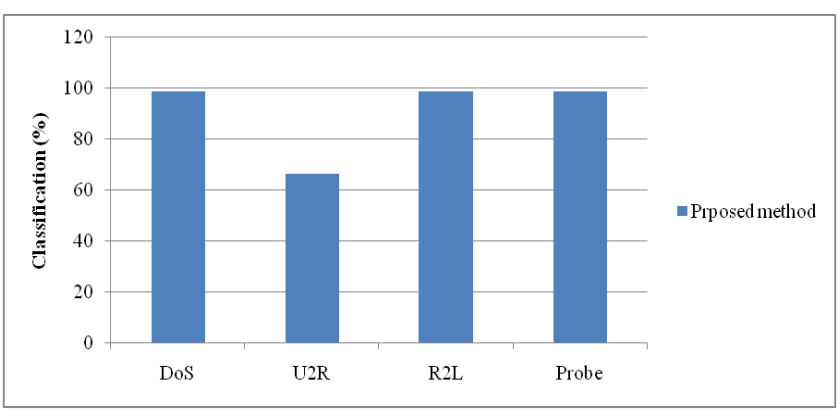

Figure3: Accuracy comparison in case of Random selection-3 classification comparison

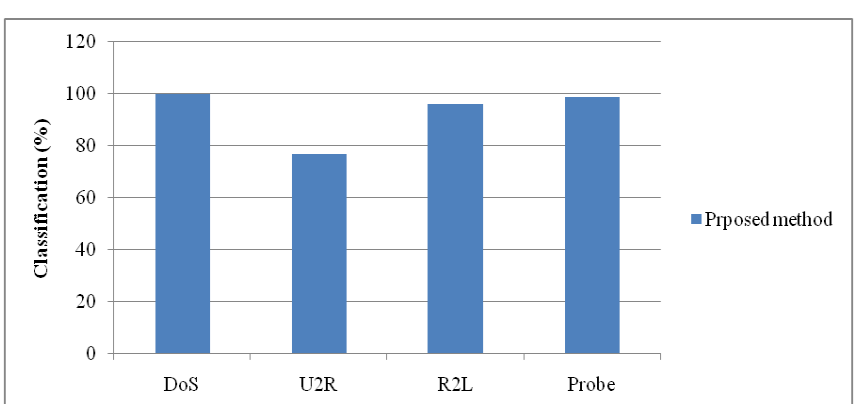

Figure4: Accuracy comparison in case of Random selection-4classification comparison

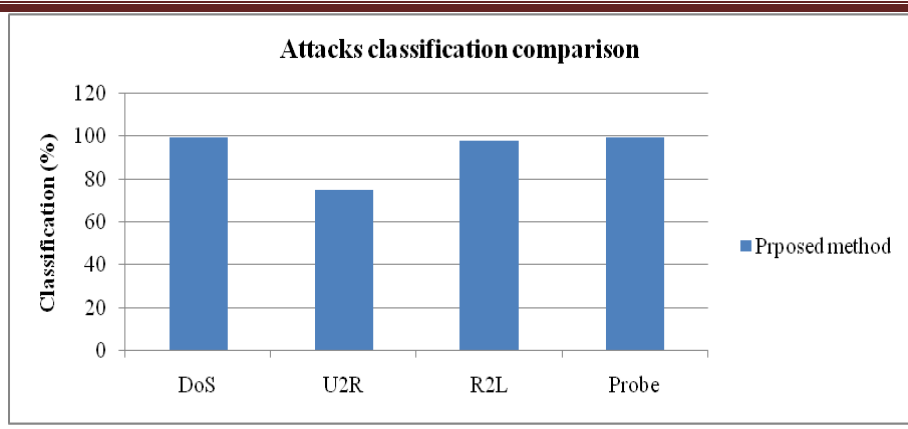

Figure 5: Average Classification Accuracy with random process

5. Conclusion

At that point we consider the ordinary information set and for discovering the intrusion. We haveascertained coordinating variable. In the first step preprocessing and grouping based on the filtration parameters. In the event that the worth crosses the farthest point esteem then the hub will be included into the last unsafe class. Then PSO have been applied on the cluster for the further data classification. The whole iteration is random so there is no biasness. It shows the results with different random selections in different perspective with different data ranges. The result shows that the classification has been improved in terms of DoS and Probe and efficient results are obtained in other cases.

References

1. Jianliang M, Haikun S, Ling B. The application on intrusion detection based on k-means cluster algorithm. In Information Technology and Applications, 2009. IFITA'09. International Forum on 2009 May 15 (Vol. 1, pp. 150-152). 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 Computer Science-Technology and Applications, 2009. IFCSTA'09. International Forum on 2009 Dec 25 (Vol. 1, pp. 76-79). 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.
6. Farhaoui Y. How to secure web servers by the intrusion prevention system (IPS)?International Journal of Advanced Computer Research. 2016 Mar 1; 6(23):65.
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.
8. Devaraju S, Ramakrishnan S. Analysis of Intrusion Detection System Using Various Neural Network classifiers. IEEE 2011. 2011:1033-8.
9. Brugger ST. Data mining methods for network intrusion detection. University of California at Davis. 2004 Jun 9.
10. Sirisha GN, Shashi M. Subspace clustering for high dimensional datasets. International Journal of Advanced Computer Research. 2016; 6(26):177.
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.
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.
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.
15. Lee W, Stolfo SJ, Mok KW. A data mining framework for building intrusion detection models. In Security and Privacy, 1999. Proceedings of the 1999 IEEE Symposium on 1999 (pp. 120-132). 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.
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 Information Technology, Computer Engineering and Management Sciences (ICM), 2011 International Conference on 2011 Sep 24 (Vol. 2, pp. 157-161). 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. In2012 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. InPattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on 2012 Mar 21 (pp. 204-209). IEEE.
22. Han LI. Using a dynamic K-means algorithm to detect anomaly activities. InComputational Intelligence and Security (CIS), 2011 Seventh International Conference on 2011 Dec 3 (pp. 1049-1052). IEEE.
23. Benaicha SE, Saoudi L, Guermeche SE, Lounis O. Intrusion detection system using genetic algorithm. InScience and Information Conference (SAI) 2014 (pp. 564-568). IEEE.
24. Thaseen IS, Kumar CA. Intrusion detection model using fusion of PCA and optimized SVM. InContemporary Computing and Informatics (IC3I), 2014 International Conference on 2014 (pp. 879-884). IEEE.
25. Wagh SK, Kolhe SR. Effective intrusion detection system using semi-supervised learning. InData Mining and Intelligent Computing (ICDMIC), 2014 International Conference on 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 Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on 2015 Feb 21 (pp. 431-436). IEEE.
28. Yan C. Intelligent Intrusion Detection Based on Soft Computing. In Measuring Technology and Mechatronics Automation (ICMTMA), 2015 Seventh International Conference on 2015 Jun 13 (pp. 577-580). IEEE.
29. Haidar GA, Boustany C. High Perception Intrusion Detection Systems Using Neural Networks.Ninth 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. InEngineering& MIS (ICEMIS), 2017 International Conference on 2017 May 8 (pp. 1-6). IEEE.
31. Ding S, Wang G. Research on intrusion detection technology based on deep learning. InComputer and Communications (ICCC), 2017 3rd IEEE International Conference on 2017 Dec 13 (pp. 1474-1478). IEEE.
32. Xiaofeng Z, Xiaohong H. Research on intrusion detection based on improved combination of K-means and multi-level SVM. InCommunication Technology (ICCT), 2017 IEEE 17th International Conference on 2017 Oct 27 (pp. 2042-2045). IEEE.
33. Potteti S, Parati N. Intrusion detection system using hybrid Fuzzy Genetic algorithm. InTrends in Electronics and Informatics (ICEI), 2017 International Conference on 2017 May 11 (pp. 613-618). 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 Mar 1;3(1):33-9.
36. Almi'ani M, Ghazleh AA, Al-Rahayfeh A, Razaque A. Intelligent intrusion detection system using clustered self-organized map. InSoftware Defined Systems (SDS), 2018 Fifth International Conference on 2018 Apr 23 (pp. 138-144). IEEE.
37. Anwer HM, Farouk M, Abdel-Hamid A. A framework for efficient network anomaly intrusion detection with features selection. InInformation and Communication Systems (ICICS), 2018 9th International Conference on 2018 Apr 3 (pp. 157-162). IEEE.