An Efficient Intrusion Detection System based on Random-Iteration Particle Swarm Optimization

Nidhi Shrivastava, Ruchi Jain, Shiv Kumar

***Abstract:** In this paper an efficient framework has been developed for efficient intrusion detection system. In the first step the data NSL-KDD cup99 is divided into k-clusters based on the filtration parameters that are content feature, traffic features and the host feature. The clusters are separated based on the support value. Then random-iteration particle swarm optimization (RI-PSO) has been applied on the cluster for the further data classification. The classification is considered for denial of service (DoS), user to root (U2R), remote to user (R2L) and probe attacks. The results are efficient in comparison to the previous methods.*

***Keywords:** Association rule mining, RIPSO, DoS, U2R, R2L, Probe*

I. INTRODUCTION

The application which is equipped for observing malevolent conduct is called intrusion detection [1]. The principle sorts of assaults considered for this review are disavowal of administration (DoS), client to root (U2R), remote to login (R2L) and test. The technique of malevolent session location, distinguishing proof when we are at present correspondence or evacuating data in the system condition [2, 3]. The key part in following and perceiving the segments and pernicious conduct is higher concern now days [3, 4]. Diverse strategies and new calculations have been proposed toward this path. Distinctive creators gave their own particular perspectives a few focal points and recommendations. So there are a few holes can be perceived in enhancing the order exactness. Interruption location structure administers dealing with the correspondence occasions occurring in PC or system conditions and separating them for indications of conceivable occasions, which are encroachment or unavoidable hazards to PC security, or standard security rehearses Intrusion identification framework (IDS) have rose to recognize practices which chance the uprightness, insurance or openness of are sourced as a push to give a reaction for existing security issues [5]. Considering the above truths we have investigated and examinations few points in the subsequent fragments.

We in like manner look at about data mining and progression strategies, in light of the way that it can be used as a piece of forming the structure which conveys a superior acknowledgment system. As we examine this review toward an unrivaled structure with the blend of information mining and streamlining. These procedures are gainful and has been utilized as a recognition approaches like [6-11]. So the utilization of these tallies can improve an effect. The investigates have expanded their viewpoints in this bearing by a couple investigation papers as in [12-15].There are a few reports are accounted for toward security and interruption recognition [7-20]. We have likewise examined the security perspectives as it can be simple for the comprehension of security dangers and their temperament.

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].

II. RELATED WORK

In 2011, Muda et al. [21] discuss the issue of current anomaly area that it not ready to perceive an extensive variety of ambushes successfully. To thrashing this issue, they propose a hybrid learning approach through blend of K-Means gathering and Naïve Bayes course of action. The proposed technique will cluster all data into the relating pack before applying a classifier for request reason. A test is finished to evaluate the execution of the proposed approach using KDD Cup '99 dataset. Result exhibits that the proposed approach performed better in term of accuracy, recognizable proof rate with sensible false ready rate.

In 2012, LI [22] displayed an enhanced FP-development calculation. They have recommended information preprocessing of which is proficient in expanding effectiveness in looking the prefix hub so that time multifaceted nature is diminished. This procedure has been connected for interruption recognition and the accomplished outcomes are viable and possible. In 2012, P. Prasenna et al. [23] suggested that in standard framework security just relies on upon logical computations and low counter measures to taken to maintain a strategic distance from interference disclosure system, though most of this approachs to the extent theoretically tried to execute. Makers suggest that rather than delivering huge number of standards the progression streamlining methodologies like hereditary system programming (GNP) can be used .The GNP relies on upon facilitated graph. They focus on the security issues related to pass on a data mining-based IDS in a steady circumstance.

Manuscript published on 30 June 2017.
***** Correspondence Author (s)

Nidhi Shrivastava, M.Tech Student, Department of CTA, Lakshmi Narain College of Technology Excellence, Bhopal (M.P)-462021, India.

Ruchi Jain, Assistant Professor, Department of Computer Science & Engineering, Lakshmi Narain College of Technology Excellence, Bhopal (M.P)-462021, India.

Dr. Shiv Kumar, Professor, Department of Computer Science & Engineering, Lakshmi Narain College of Technology Excellence, Bhopal (M.P)-462021, India.

They aggregate up the issue of GNP with connection rule mining and propose a feathery weighted association guideline mining with GNP framework fitting for both constant and discrete attributes. In 2014, Deshmukh et al. [24]. Presents an information mining procedure in which distinctive preprocessing systems are incorporated, for instance, Normalization, Discretization and Feature assurance. With the help of these systems the data is preprocessed and required components are picked. They used NaIve Bayes method as a piece of a managed learning technique which bunches distinctive framework events for the KDD cup'99 Dataset.

In 2014, Benaicha et al. [25] exhibit a Genetic Algorithm (GA) approach with an improved beginning people and decision overseer, to gainfully recognize diverse sorts of framework intrusions. They used GA to redesign the quest for ambush circumstances in audit records, as a result of its incredible equality examination/manhandle; as demonstrated by the makers it gives the subset of potential strikes which are accessible in the survey archive in a sensible planning time. The testing time of the Network Security Laboratory Knowledge Discovery and Data Mining (NSL-KDD99) benchmark dataset has been used to recognize the manhandle works out. Their approach of IDS with Genetic computation grows the execution of the area rate of the Network Intrusion Detection Model and reductions the false positive rate.

In 2014 Kiss et al. [26] suggest that Modern Networked Critical Infrastructures (NCI), including advanced and physical structures, is exhibited to shrewd computerized attacks concentrating on the enduring operation of these systems. To ensure irregularity care, their watched data can be used as a piece of comprehension with data mining techniques to make Intrusion Detection Systems (IDS) or Anomaly Detection Systems (ADS). They proposed a batching based philosophy for recognizing computerized strikes that realize peculiarities in NCI. Distinctive clustering techniques are examined to pick the most fitting for gathering the time-course of action data highlights, in this manner requesting the states and potential advanced ambushes to the physical structure. The Hadoop utilization of MapReduce perspective is used to give a fitting taking care of condition to huge datasets.

In 2014, Thaseen et al. [27] proposed a novel technique for fusing basic section examination (PCA) and reinforce vector machine (SVM) by enhancing the bit parameters using customized parameter assurance strategy. Their philosophy diminishes the arrangement and testing time to recognize intrusions in this way upgrading the exactness. Their proposed system was attempted on KDD data set. The datasets were intentionally separated into get ready and testing considering the minority strikes, for instance, U2R and R2L to be accessible in the testing set to perceive the occasion of cloud attack. Their results show that the proposed methodology is productive in recognizing interferences. Their test outcomes exhibit that the gathering precision of the proposed system beats other request techniques using SVM as the classifier and other dimensionality reducing or highlight assurance methodologies.

In 2014, Wagh et al. [28] suggested Network security is an imperative piece of web enabled structures in the present world circumstance. As demonstrated by the makers due to mind boggling chain of PCs the open entryways for intrusions and attacks have extended. Appropriately it is need of incredible significance to find the best courses possible to secure our systems. So the makers propose interference area structures are expecting fundamental part for PC security. The best procedure used to deal with issue of IDS is machine learning. Thy watched that the rising field of semi directed learning offers an ensured course for correlative investigation. So they proposed a semi-guided procedure to reduce false ready rate and to upgrade area rate for IDS.

In 2014, Masarat et al. [29] introduced a novel multistep structure in perspective of machine learning frameworks to make a successful classifier. In initial step, the part decision technique will complete considering get extent of components by the makers. Their strategy can upgrade the execution of classifiers which are made considering these segments. In classifiers mix step, we will present a novel cushioned assembling technique. In this manner, classifiers with more execution and lower cost have more effect to make the last classifier.

In 2015, Bahl et al. [30] recommended U2R assault class is an open research issue. Their motivation of this review is to distinguish the imperative components to enhance the location rate and decrease the false recognition rate. The explored highlight subset decision strategies improve the general precision, disclosure rate of U2R assault class moreover diminish the computational cost. The trial comes about have exhibited a recognizable change in area rate of U2R assault class with highlight subset assurance frameworks.

In 2015, Yan et al. [31] proposed a smart interruption identification demonstrate. Considering the traits of overall transcendence of inherited figuring and territory of nerve, the model redesigns the weights of the neural framework using hereditary calculation. Test comes about exhibit that the wise way can improve the capability of the interference ID.

In 2015, Haidar et al. [32] accentuates the hugeness of variation from the norm based intrusion acknowledgment systems, the indispensable aftereffects of these structures, latest made procedures and what is typical from the future trials in this field. Furthermore, the technique for learning customer profiles impacts in perceiving intrusions can be investigated. At long last, the lights will be shed on a disconnected approach utilizing Multi-Layer Perceptron (MLP) and Self Organizing Maps (SOM) which is a recognized strategy in interruption identification.

III. 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.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

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. The data is classified either individual or in the group. In our methodology NSL-KDD dataset has been considered.

It is information set which does exclude repetitive record and test sets. At that point equivalent extent of approx. 10,000 records from the entire dataset has been selected. It is group into two parts first is safe and another is unsafe.

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 k-clusters based on the filtration parameters that are content feature, traffic features and the host feature. The clusters are separated based on the support value. 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.

Data separation will take place on the selected database. The data separation will be done based on the fourth field and categorized according to the connection value that is safe or unsafe. First safe nodes are considered with normal receive and other as the unsafe receive. Then we again filter the attack data based on the receiving connection as the normal and prepare the initial attack data.

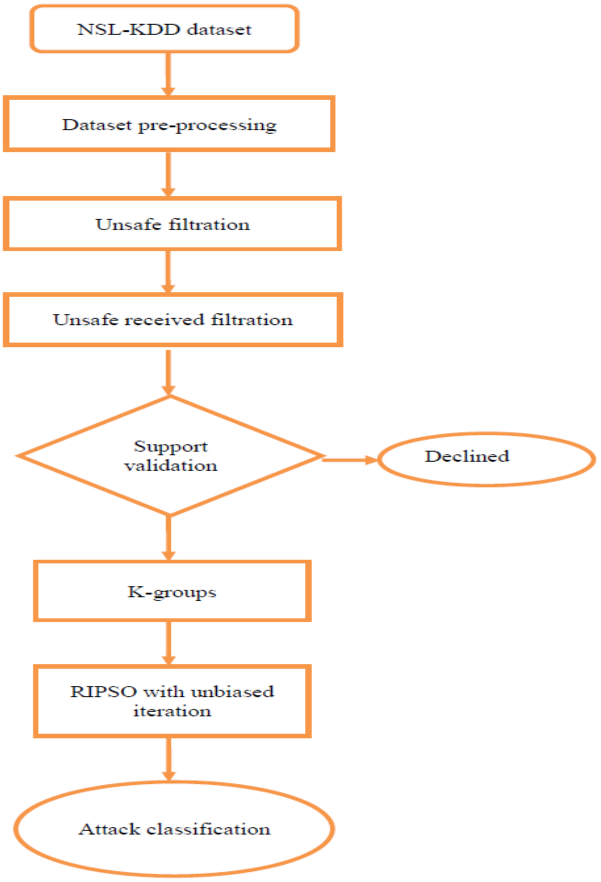

Figure 1: Flowchart of my Proposed Technique

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

RIPSO 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=∑psois1* Rv +psois2* Rv +psois3* Rv ….psois5* Rv /n
for 2 to 5 do
Ns2=Kw+∑psois1u * Rv +psois2u * Rv +psois3u * Rv ….psois5u * Rv /n - Rvp
Ns3= Ns2+∑ acois1u * Rv +acois2u * Rv +acois3u * Rv ….acois5u * Rv /n- Rvp
Ns4= Ns3+∑ acois1u * Rv +acois2u * Rv +acois3u * Rv ….acois5u * Rv /n- Rvp
Ns5= Ns4+∑ 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.
```

IV. RESULTS

In this section the results obtained from our method has been discussed. The results are prepared based on the nodes which are safe but not received as the safe IDs. Table 1 shows the results with different random selections in different perspective with different data ranges. It shows the classification has been improved in terms of DoS and Probe and efficient results are obtained in other cases. Figure 2 shows the graphical representation of achieved classification accuracy. Figure 3 shows the overall comparison from different previous methods.

Table 1: Random Node Selection and Classification Comparison

Details	Values	DoS (%)	U2R (%)	R2L (%)	Probe (%)
ID: 66620-70252					
Normal	2182	100	75	96.15	100
Attack	1451				
Unsafe	1333				
Filter	340				
Final Set	1673				
ID: 66622-76312					
Normal	5830	100	83.33	98.50	100
Attack	3861				
Unsafe	3553				
Filter	861				
Final Set	4414				
ID: 98102-106430					
Normal	4986	100	66.66	98.66	100
Attack	3343				
Unsafe	3100				
Filter	782				
Final Set	3882				
ID: 90732-102733					
Normal	7206	100	75	98.91	100
Attack	4796				
Unsafe	4405				
Filter	1123				
Final Set	5528				
Average accuracy		100	74.75	97.5	100

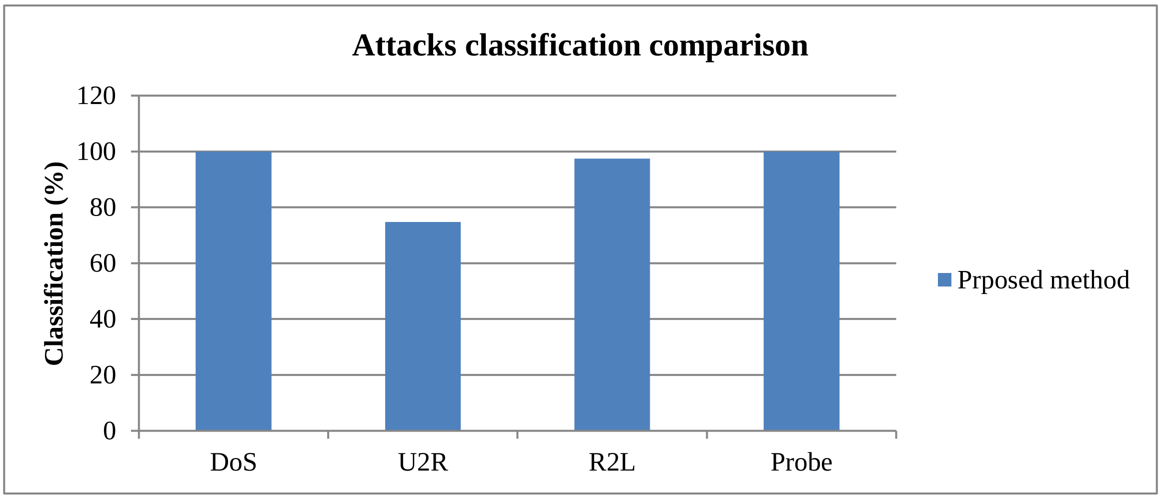

Figure 2 Average Classification Accuracy with Random Process

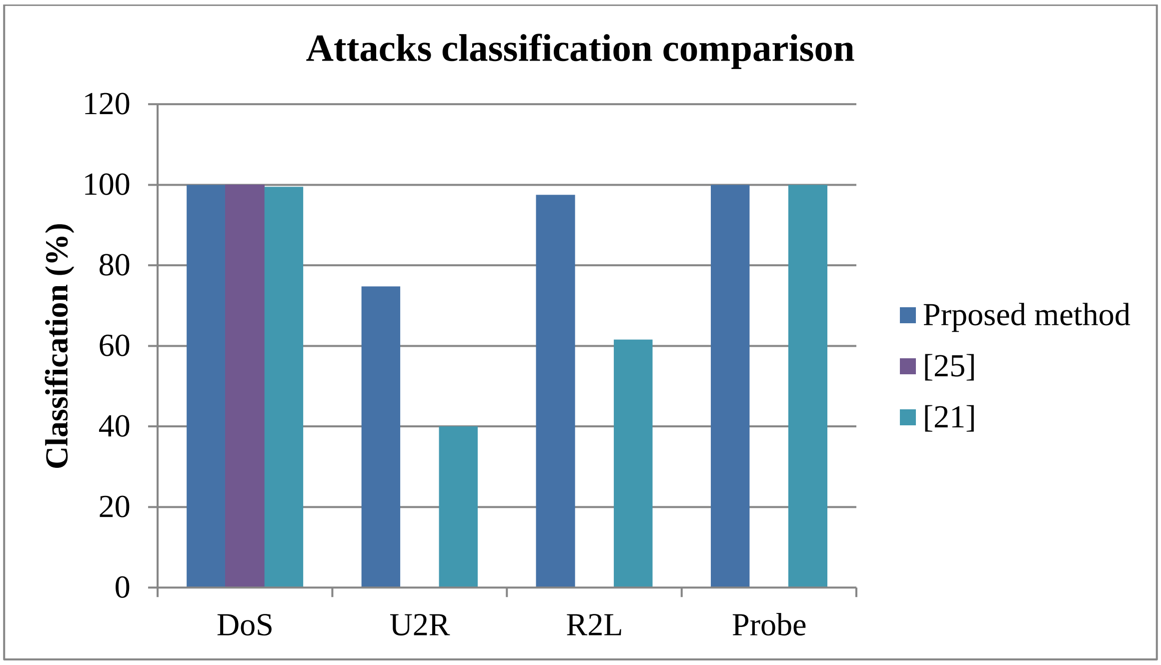

Figure 3 Classification Comparisons with Previous Method with Random Process

V. CONCLUSION

In our methodology NSL-KDD dataset has been considered. It is information set which does exclude repetitive record and a test set it is group into two parts first is safe and another is unsafe. 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 k-clusters based on the filtration parameters that are content feature, traffic features and the host feature. The clusters are separated based on the support value. In the event that the worth crosses the farthest point esteem then the hub will be included into the last unsafe class. Then RIPSO 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. 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.
2. 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.
3. Kabiri P, Ghorbani AA. Research on Intrusion Detection and Response: A Survey. IJ Network Security. 2005 Sep; 1(2):84-102.
4. Park HA. Secure chip based encrypted search protocol in mobile office environments. International Journal of Advanced Computer Research. 2016; 6(24):72-80.
5. 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.
6. 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.
7. Devaraju S, Ramakrishnan S. Analysis of Intrusion Detection System Using Various Neural Network classifiers. IEEE 2011. 2011:1033-8.
8. Conteh NY, Schmick PJ. Cybersecurity: risks, vulnerabilities and countermeasures to prevent social engineering attacks. International Journal of Advanced Computer Research. 2016 Mar 1; 6(23):31.
9. Lee HY, Wang NJ. The implementation and investigation of securing web applications upon multi-platform for a single sign-on functionality. International Journal of Advanced Computer Research. 2016 Mar 1; 6(23):39.
10. Ishida M, Takakura H, Okabe Y. High-performance intrusion detection using optigrid clustering and grid-based labelling. InApplications and the Internet (SAINT), 2011 IEEE/IPSJ 11th International Symposium on 2011 Jul 18 (pp. 11-19). IEEE.
11. Brugger ST. Data mining methods for network intrusion detection. University of California at Davis. 2004 Jun 9.
12. Lee W, Stolfo SJ. Data mining approaches for intrusion detection. In Usenix security 1998 Jan 26.
13. Nalavade K, Meshram BB. Mining Association Rules to Evade Network Intrusion in Network Audit Data. International Journal of Advanced Computer Research. 2014 Jun 1;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 Jun 1; 4(2):487.
15. Lee W, Stolfo SJ, Mok KW. A data mining framework for building intrusion detection models. InSecurity and Privacy, 1999. Proceedings of the 1999 IEEE Symposium on 1999 (pp. 120-132). IEEE.
16. Kumari S, Shrivastava M. A Study Paper on IDS Attack Classification Using Various Data Mining Techniques. International Journal of Advanced Computer Research. 2012; 2(3).
17. Venkatesan R, Ganesan R, Selvakumar AA. A Comprehensive Study in Data Mining Frameworks for Intrusion Detection. International Journal of Advanced Computer Research (IJACR). 2012; 2: 29-34.
18. Patel R, Bakhshi D, Arjariya T. Random Particle Swarm Optimization (RPSO) based Intrusion Detection System. International Journal of Advanced. 2015; 2(5): 60-66.
19. Sperotto A, Schaffrath G, Sadre R, Morariu C, Pras A, Stiller B. An overview of IP flow-based intrusion detection. Communications Surveys & Tutorials, IEEE. 2010 Jul 1; 12(3):343-56.
20. Han LI. Using a dynamic K-means algorithm to detect anomaly activities. In Computational Intelligence and Security (CIS), 2011 Seventh International Conference on 2011 Dec 3 (pp. 1049-1052). IEEE.
21. Muda Z, Yassin W, Sulaiman MN, Udzir NI. Intrusion detection based on K-Means clustering and Naïve Bayes classification. In Information Technology in Asia (CITA 11), 2011 7th International Conference on 2011 Jul 12 (pp. 1-6). IEEE.

22. Yin-huan LI. Design of intrusion detection model based on data mining technology. In2012 International Conference on Industrial Control and Electronics Engineering 2012 Aug 23.
23. Prasenna P, RaghavRamana AV, Krishnakumar R, 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.
24. Deshmukh DH, Ghorpade T, Padiya P. Intrusion detection system by improved preprocessing methods and Naïve Bayes classifier using NSL-KDD 99 Dataset. In Electronics and Communication Systems (ICECS), 2014 International Conference on 2014 Feb 13 (pp. 1-7). IEEE.
25. Benaicha SE, Saoudi L, Guermeche B, Eddine S, Lounis O. Intrusion detection system using genetic algorithm. InScience and Information Conference (SAI), 2014 2014 Aug 27 (pp. 564-568). IEEE.
26. Kiss I, Genge B, Haller P, Sebestyen G. Data clustering-based anomaly detection in industrial control systems. In Intelligent Computer Communication and Processing (ICCP), 2014 IEEE International Conference on 2014 Sep 4 (pp. 275-281). IEEE.
27. Thaseen IS, Kumar CA. Intrusion detection model using fusion of PCA and optimized SVM. In Contemporary Computing and Informatics (IC3I), 2014 International Conference on 2014 Nov 27 (pp. 879-884). IEEE.
28. Wagh SK, Kolhe SR. Effective intrusion detection system using semi-supervised learning. In Data Mining and Intelligent Computing (ICDMIC), 2014 International Conference on 2014 Sep 5 (pp. 1-5). IEEE.
29. Masarat S, Taheri H, Sharifian S. A novel framework based on fuzzy ensemble of classifiers for intrusion detection systems. In Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on 2014 Oct 29 (pp. 165-170). IEEE.
30. 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.
31. 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.
32. 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.