International Journal of Advanced Technology and Engineering Exploration, Vol 3(20)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/ 10.19101/IJATEE.2016.320005

Review Article

An approach for efficient intrusion detection based on R-ACO

Namita Sharma[1*] and Bhupesh Gaur[2]
M.Tech Research Scholar, Computer Science, Technocrat Institute of Technology, Bhopal, India[1]
Professor & Head, Computer Science, Technocrat Institute of Technology, Bhopal, India[2]

Abstract
Our paper focuses on the limitation faced in the traditional approaches. In this paper a hybrid framework based on associated clusters and random ant colony optimization (R-ACO). In our approach the dataset of NSL-KDD have been considered. It is a data set which does not include redundant record and test sets. Then equal proportion dataset from the whole dataset are selected. The data is pre-processed according to the normal data filtration and attack data filtration. Then normal data based on the intrusion filed is pre- processed which are not received as the normal set. This dataset is passed for k1-k6 transaction for finding the associated cluster based on the property. Then R-ACO for finding the global optimum value has been applied. If the optimum value satisfied the threshold, then the node will be added into the final attack category. Finally based on the attack category of Denial of Service (DoS), User to Root (U2R), Remote to User (R2L) and Probing (Probe) based on the final classification. Our results support better classification in comparison to the previous techniques used in several research papers as per our study.

Keywords
Intrusion detection, R-ACO, DOS, U2R, R2L, Probe.

1.Introduction
Starting late, various experts are focused to use data burrowing thoughts for Intrusion Detection [1]. This is a system to remove the comprehended information and learning. Interruption location is the strategy of malignant on the structure and framework when we are as of now correspondence or expelling data in the consistent environment [2][3]. Since its creation, intrusion area has been one of the key segments in achieving information security. It goes about as the second-line monitor, which supplements the identification controls. Exactly when the controls failed, the interference recognizable proof structures should have the ability to distinguish it consistent and alert the security officers to take prompt and appropriate exercises [3][4]. Interference recognition structure oversees managing the events happening in PC structure or framework circumstances and dissecting them for signs of possible events, which are infringement or unavoidable risks to PC security, or standard security practices Intrusion discovery framework (IDS) have ascended to distinguish exercises which imperil the uprightness, protection or openness of are sourced as a push to give a response for existing security issues [5].

*Author for correspondence

So in the above course we audit a couple points in the subsequent portions. We in like manner inspect about data mining and headway techniques, in light of the way that it can be used as the structure which conveys a superior acknowledgment system. As we inspect this study toward a prevalent framework with the mix of data mining and streamlining. These systems are important and has been used as a piece of differing approaches like [6][7][8][9][10][11]. So the usage of these counts can enhance an impact. The scrutinizes have expanded their perspectives in this bearing by a few examination papers as in [12][13][14][15].

In this paper an efficient method based on ACO has been presented.

2.Literature survey
In 2010, G. Schaffrath et al. [16] give a study of ebb and flow research in the zone of stream based interruption recognition. The study begins with an inspiration why stream based interruption identification is required. The idea of streams is clarified, and applicable principles are recognized. The paper gives a characterization of assaults and safeguard methods and shows how stream based strategies can be utilized to distinguish filters, worms, Botnets and DoS assaults. In 2011, Zhengjie Li et al.

[17] propose a K-implies grouping calculation in light of molecule swarm advancement (PSO-KM). The proposed calculation has overcome falling into neighbourhood minima and has moderately great general merged. Probes information sets KDD CUP 99 have demonstrated the viability of the proposed technique furthermore demonstrates the strategy has higher recognition rate and lower false location rate. In 2012, LI Yin–huan [18] concentrates on an enhanced FP-Growth calculation. As indicated by creator pre-processing of information mining can expand effectiveness on seeking the basic prefix of hub and decrease the time multifaceted nature of building FP-tree. Taking into account the enhanced FP Growth calculation and other information mining systems, an interruption location model is completed by creators. Their test results are successful and plausible. In 2012, P. Prasenna et al. [19] proposed that in customary system security essentially depends on scientific calculations and low counter measures to taken to avoid interruption identification framework, albeit a large portion of this methodologies as far as hypothetically tested to execute. Creators recommend that as opposed to producing substantial number of tenets the advancement streamlining procedures like Genetic Network Programming (GNP) can be utilized. In 2011, LI Han [20] concentrates on interruption discovery in view of bunching examination. The point is to enhance the discovery rate and decline the false alert rate. An altered element K-implies calculation called MDKM to identify irregularity exercises is proposed and relating reproduction investigations are displayed. Firstly, the MDKM calculation channels the commotion and disengaged focuses on the information set. In 2011, Z. Muda et al. [21] examine about the issue of current irregularity recognition that it not able to recognize a wide range of assaults accurately. To defeat this issue, they propose a mixture learning approach through blend of K-Means bunching and Naïve Bayes grouping.

The proposed methodology will bunch all information into the relating bunch before applying a classifier for arrangement reason. An investigation is completed to assess the execution of the proposed approach utilizing KDD Cup '99 dataset. Results demonstrate that the proposed approach performed better in term of exactness, identification rate with sensible false caution rate. In 2014, Deshmukh et al. [22] presents a Data Mining technique in which different pre-processing strategies are included, for example, Normalization, Discretization and Feature choice. With the assistance of these techniques the information is pre-processed and required elements are chosen. They utilized Naive Bayes strategy as a part of a managed learning technique which orders different system occasions for the KDD cup'99 Dataset. In 2014, Benaicha et al. [23] present a Genetic Algorithm (GA) approach with an enhanced starting populace and determination administrator, to proficiently recognize different sorts of system interruptions. They utilized GA to streamline the pursuit of assault situations in review documents, on account of its great parity investigation/abuse; as indicated by the creators it gives the subset of potential assaults which are available in the review record in a sensible handling time.

The testing period of the Network Security Laboratory Knowledge Discovery and Data Mining (NSL-KDD99) benchmark dataset has been utilized to identify the abuse exercises. Their methodology of IDS with Genetic calculation expands the execution of the discovery rate of the Network Intrusion Detection Model and lessens the false positive rate. In 2014 Kiss et al. [24] propose that Modern Networked Critical Infrastructures (NCI), including digital and physical frameworks, is presented to clever digital assaults focusing on the steady operation of these frameworks. To guarantee irregularity mindfulness, their watched information can be utilized as a part of agreement with information mining methods to create Intrusion Detection Systems (IDS) or Anomaly Detection Systems (ADS). They proposed a bunching based methodology for identifying digital assaults that cause peculiarities in NCI. Different grouping methods are investigated to pick the most appropriate for bunching the time-arrangement information highlights, along these lines characterizing the states and potential digital assaults to the physical framework. The Hadoop usage of MapReduce worldview is utilized to give an appropriate preparing environment to huge datasets. In 2014, Thaseen et al. [25] proposed a novel strategy for incorporating essential segment examination (PCA) and bolster vector machine (SVM) by advancing the piece parameters utilizing programmed parameter determination method. Their methodology decreases the preparation and testing time to recognize interruptions in this way enhancing the precision. In 2014, Wagh et al. [26] recommended Network security is a vital part of web empowered frameworks in the present world situation. As per the creators because of many-sided 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 courses conceivable to ensure our frameworks. So the creators recommend interruption identification frameworks are assuming basic part for PC security.

In 2014, Masarat et al. [27] presented a novel multistep structure in view of machine learning systems to make a productive classifier. In initial step, the component determination strategy will execute in view of increase proportion of elements by the creators. Their technique can enhance the execution of classifiers which are made taking into account these components. In classifiers mix step, we will show a novel fluffy gathering strategy. In this way, classifiers with more execution and lower cost have more impact to make the last classifier.

3.proposed work

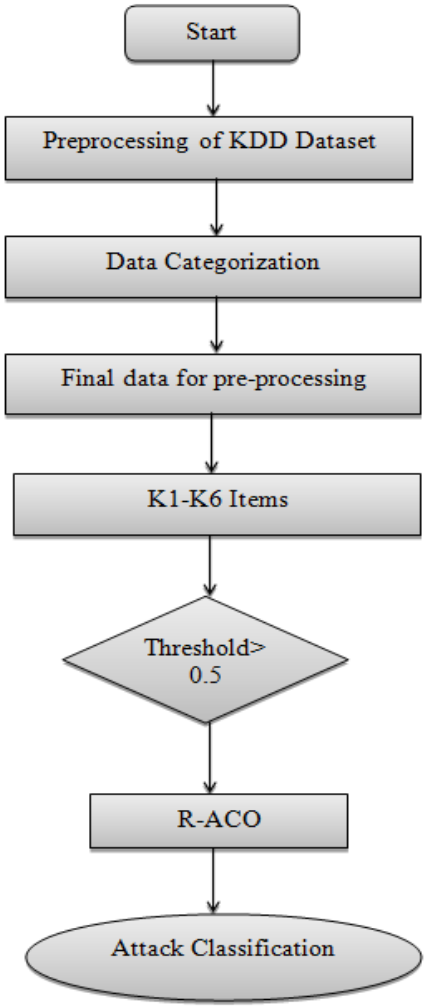

Figure 1 Flowchart

In our approach we have considered the dataset of NSL-KDD. It is a data set which does not include redundant record and test sets. Then we consider equal proportion dataset from the whole dataset. The data is preprocessed according to the normal data filtration and attack data filtration. Then normal data based on the intrusion filed is pre- processed which are not received as the normal set. This dataset is passed for k1-k6 transaction for finding the associated cluster based on the property. Then we apply R-ACO for finding the global optimum value. If the optimum value satisfied the threshold, then the node will be added into the final attack category. Finally based on the attack category of Denial of Service (DoS), User to Root (U2R), Remote to User (R2L) and Probing (Probe) based on the final classification. Our results support better classification in comparison to the previous techniques used in several research papers as per our study. *Figure 1* shows the flowchart of the proposed technique.

This approach is divided into five different parts as shown below.

1)Pre-processing
Data pre-processing is applied on the whole database. Data pre-processing is a means of selecting any random records from 1025973 records. The data proportion is taken in this way that the chances of getting all the attacks are possible.

2)Pre-processed data and Associated K-items
Then we consider equal proportion of data from the whole dataset. The data is pre-processed according to the normal data filtration and attack data filtration. Then normal data based on the intrusion filed is pre-processed which are not received as the normal set. The associated k-items are pre-processed for creating the final set which can apply for the optimization process.

3)Random Ant Colony Optimization (R-ACO)
Then we apply R-ACO for the better classification. The algorithm is shown below:
Input:
•Ant-System(as1,as2….asn)
Output:
•Final optimized system (fos1,fos2…….fosn)
Terminology used in this algorithm:
as: ant system
asp: previous ant system
gr: generated random value.
ev: evaporation value
ft: first trail

st: second trail
tt: third trail
i: variable
n: total ant
Step 1: Data selection
Step 2: define as an ant system
Step 3: Pre-processed data and Associated K-items
Step 4: Initialize the ant system
Step 5: for i=1to 3 trails
ev_i = random value between(gr(0,1))
ft_1 = (as1+ as2+as3+as4….+asn)/n
$ft_i = ft_1$+ (as1+ as2+as3+as4….+asn)/n- ev_i
Check the threshold value
if ($as_i > asp_i$)
$asp_i = as_i$
else
$asp_i = asp_i$
Step 6: Check the final classification.
Step 7: It is based on the whole ant system

Total=$\sum as_i$ / n
Step 8: If it is greater than the threshold value then it is classifies as the attack value.
Step 9: Finish

The above algorithm clearly shows the working phenomena based on associated clusters and R-ACOP.

4)Attack Classification
This classification is based on the attacks property. It is match for the classification.

5)Final Analysis
Last investigation is done on the premise of contrasting the last attack database and the aggregate database. It will be better clarified in our outcome investigation. The outcome demonstrates the better characterization as far as DoS and Probe.

4.Result analysis

In this section the results obtained from our method has been discussed. The results are prepared based on the nodes which are not obtained as the normal node. It is shown in *Table 1*. These values are then applied to find the optimal threshold by applying R-ACO method. It is filtered based on the fifty percentage threshold value. If the value is greater than the received highest threshold value then it qualifies otherwise it not qualifies. Based on the obtained value the final attack database is created. The results are shown in *Figure 2 to Figure 4.* The comparison from the previous results is shown in *Figure 5*. It shows the significant improvements have been obtained from the previous approach.

Table 1 Classified clusters

Node	K1	K2	K3	K4	K5	K6
98111	0.8462	0.9231	0.4444	0.6667	0.6	0.5
98125	1	1	0.4444	0.6667	0.5	0.5
98153	1	1	0.4444	0.5556	0.7	0.5
98154	1	1	0.2222	0.6667	0.3	0.6
98158	1	1	0.3333	0.6667	0.6	0.5
98190	1	1	0.3333	0.6667	0.3	0.6
98200	1	1	0.3333	0.6667	0.4	0.7

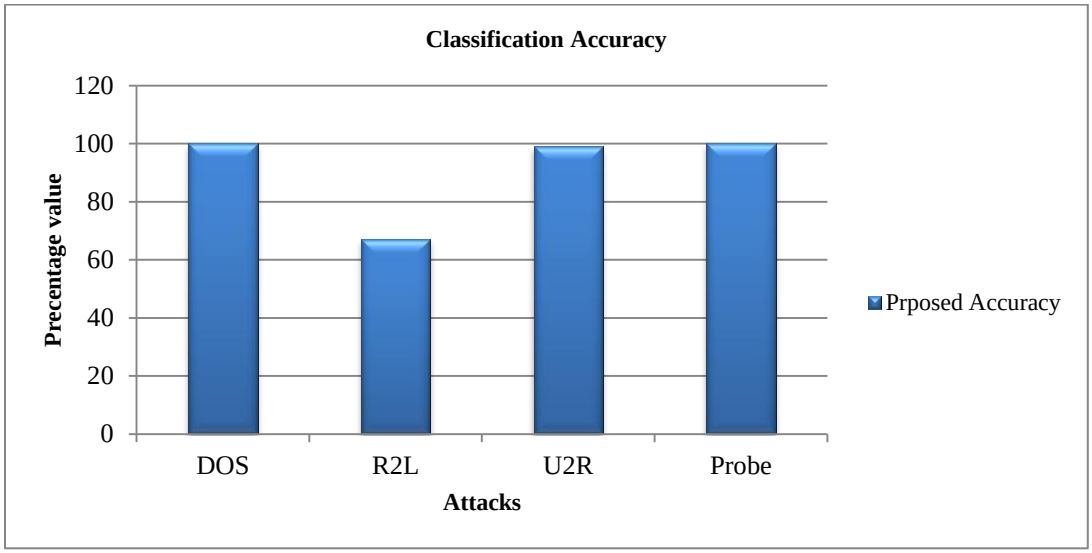

Figure 2 Attack classification accuracy (node: 98102-106540)

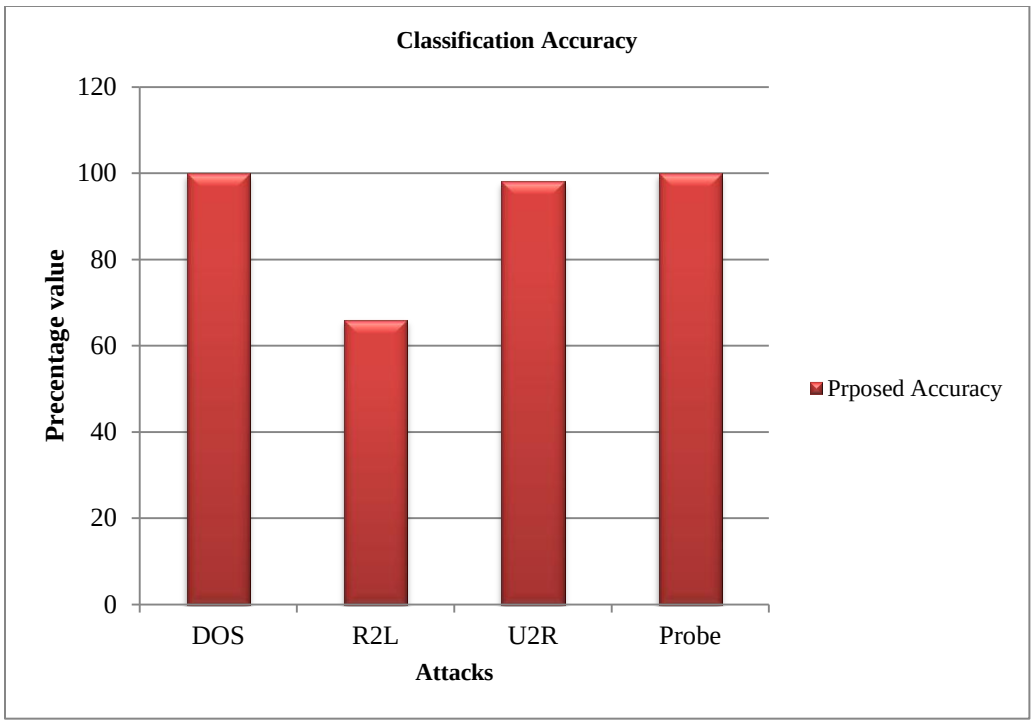

Figure 3 Attack classification accuracy (Node: 97112-104377)

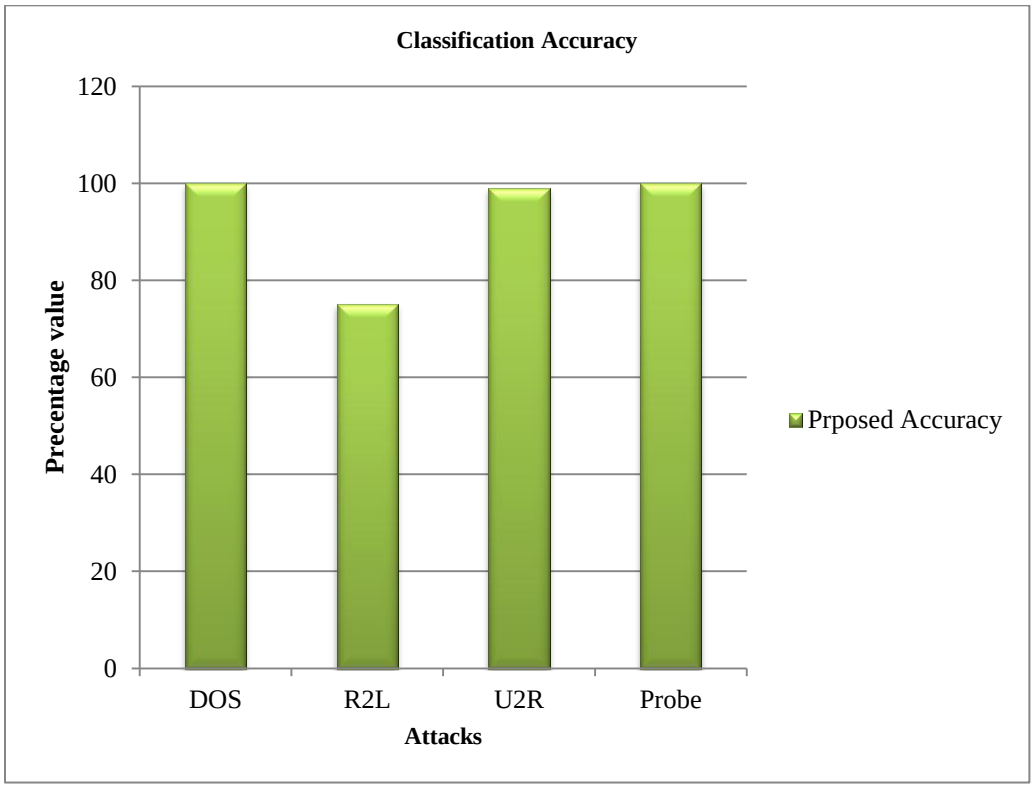

Figure 4 Attack classification accuracy (Node: 90742-102743)

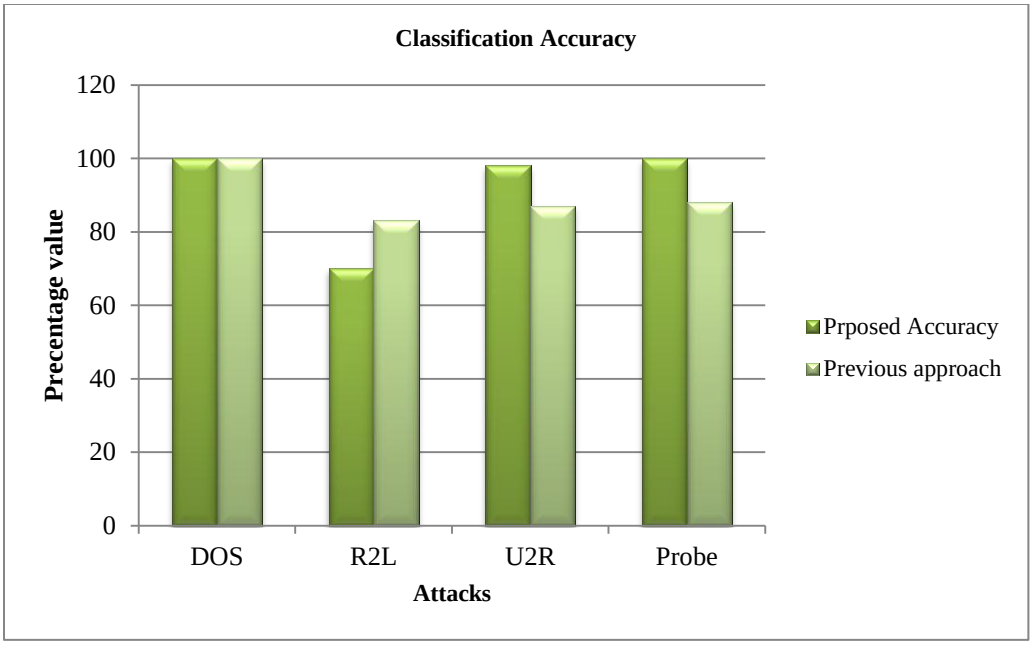

Figure 5 Accuracy comparison

5.Conclusion

In this paper random evaporation value based ACO(R-ACO) has been applied which is able to calculate the optimum threshold from the associated clusters. This classification is applied on the nodes which are not received normal and term as the possible malicious node. Our identification is based on the fixed threshold means the higher thresholds then the fixed thresholds are classified and categorized as the attack node otherwise it is a normal node. Our approach is capable for classifying the possible four attacks name DoS, U2R, R2L and Probe. The results show the classification has been improved in terms of DoS and Probe and efficient results are obtained in other cases.

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] Tavallaee M, Stakhanova N, Ghorbani AA. Toward credible evaluation of anomaly-based intrusion-detection methods. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2010; 40(5): 516-24.

[3] 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.

[4] 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.

[5] Ishida M, Takakura H, Okabe Y. High-performance intrusion detection using optigrid clustering and grid-based labelling. In 11th international symposium on applications and the internet 2011 (pp. 11-9). IEEE.

[6] Brugger ST. Data mining methods for network intrusion detection. University of California at Davis. 2004.

[7] Kailashiya D, Jain RC. Improve intrusion detection using decision tree with sampling. International Journal of Computer Technology and Applications. 2012;3(3): 1209-16.

[8] 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.

[9] Lee W, Stolfo SJ. Data mining approaches for intrusion detection. In Usenix security 1998.

[10] 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.

[11] 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.

[12] Kumari S, Shrivastava M. A study paper on IDS attack classification using various data mining techniques. International Journal of Advanced Computer Research. 2012; 2(5):195-200.
[13] Venkatesan R, Ganesan R, Selvakumar AA. A comprehensive study in data mining frameworks for intrusion detection. International Journal of Advanced Computer Research. 2012; 2(7):29-34.
[14] Farhaoui Y. How to secure web servers by the intrusion prevention system (IPS)? International Journal of Advanced Computer Research. 2016; 6(23):65-71.
[15] Patel R, Bakhshi D, Arjariya T. Random particle swarm optimization (RPSO) based intrusion detection system. International Journal of Advanced Technology and Engineering Exploration (IJATEE). 2015; 2(5):60-6.
[16] 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.
[17] 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.
[18] Yin-huan LI. Design of intrusion detection model based on data mining technology. In international conference on industrial control and electronics engineering 2012.
[19] 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.
[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 (pp. 1049-52). IEEE.
[21] 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.
[22] 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 (ICECS) 2014 (pp. 1-7). IEEE.
[23] Benaicha SE, Saoudi L, Guermeche SE, Lounis O. Intrusion detection system using genetic algorithm. In science and information conference 2014 (pp. 564-8). IEEE.
[24] Kiss I, Genge B, Haller P, Sebestyén G. Data clustering-based anomaly detection in industrial control systems. In international conference on intelligent computer communication and processing 2014 (pp. 275-81). IEEE.
[25] 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.
[26] 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.
[27] Masarat S, Taheri H, Sharifian S. A novel framework based on fuzzy ensemble of classifiers for intrusion detection systems. In international econference on computer and knowledge engineering 2014 (pp. 165-70). IEEE.