Volume 5, Issue 5, May 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Network Intrusion Detection Using Feature Selection and PROAFTN Classification

Anoop Kumar Saroj
PG-Scholar, PCST
Bhopal, M.P. (India)

Prof. Parmalik Pauranik
Asstt. Professor, PCST
Bhopal, M.P. (India)

Prof. Sandeep Kumar Singh
Asstt. Professor, PCST
Bhopal, M.P. (India)

Abstract: Classification and detection of network based intrusion is very critical task. The processing of classification and detection faced a problem of large number of attribute and mixed category of data. Day to day increases diversity of attacker and hacker generates new pattern of file for attack purpose, the process of classification and detection suffered due to this reason. The process of classification of PROAFTN is a combination of fuzzy logic and protein cell classification technique. In PROAFTN classification process all features come to the predefined classes for the classification. Now the process of improvement need some important feature selection process for increasing the classification ratio and process of classification.The particle of swarm optimization is dynamic population based searching technique. In the searching technique of particle of swarm optimization select optimal feature set for the classification process of PROAFTN classification process. The process of optimal selection of feature set increase the classification and detection ratio of modified PROAFTN classification process.For the evaluation of performance of modified PROAFTN classification technique used MATLAB 7.8.0 software. MATLAB is well known algorithm analysis software. For the analysis of PROAFTN classification process used fuzzy set tools and some standard tools of MATLAB.For the processing of input data used KDDCUP99 dataset. KDDCUP99 dataset is well known dataset for the purpose of network based intrusion detection and classification. Our classification and detection ratio in some attack case achieved 100%

Keywords: - intrusion detection, classification, KDDCUP99

I. INTRODUCTION

The classification and detection of network intrusion detection is very challenging task due to large spectrum of hacker and attacker. Day to day come into new format and dynamic nature based attack pattern for computer and host. The countermeasure of intrusion attack classification is very critical task. For the classification and detection purpose used various data mining and machine learning technique for network based intrusion detection [1,2]. The process of machine learning is very efficient for the processing of network based intrusion detection technique. Some authors also used soft computing approach with data mining and machine learning technique. The soft computing approach provide the great versatility of fuzzy logic. Fuzzy logic is process of multi-decision support system used for the prediction of rule generation and formation [5, 6]. The generation of rule and formation of rule decide the selection process of attribute in intrusion detection process. The processing of file in network is very complex, basically it contains three type of data one is traffic data second one is header data and finally it contains basic feature of network file system. All these attribute collected in single point for the process of classification and detection. The distantness of feature attribute is very complex task so some authors used feature reduction cum classification of intrusion detection technique. In this dissertation modified the PROAFTN classification technique using particle of swarm optimization for feature selection. Here PROAFTN network are used for the classification process. The selection of feature used particle of swarm optimization technique. In the continuity of this chapter discuss PROAFTN classification technique, particle of swarm optimization, feature selection and reduction, classification process, proposed algorithm and finally discuss proposed mode for network based intrusion detection technique.Feature selection is play an important role in PROAFTN classification technique [11.12]. The selection of feature gives the better classification ratio for network based intrusion detection technique. The selection of feature basically compromised with three types of data. The all process data are in different level. The selection of feature compromised with particle of swarm optimization technique. For example features that can distinguish certain type of traffic from the traffic flows are picked for the network traffic model training. The idea behind the feature selection tools is to reduce the amount of features into a feasible subset of features that do not correlate with each other.In Particle Swarm Optimization [10] optimizes an objective function by undertaking a population based search. The population comprise of possible solutions, named particles, which are metaphor of birds in flocks. These particles are at random initialized and freely fly across the multi-dimensional seek space. During flight, each particle updates its own velocity and position based on the best experience of its own and the entire population. The rest of paper describe in section II PROFTAN classification. In section III discuss particle of swarm optimization. In section IV discuss proposed algorithm and model. In section V discuss experimental results and finally conclude in section VI.

II. PROFTAN CLASSIFICATION

In this section discuss PROAFTN classification technique, these technique basically based on fuzzy logic technique and protein based classification technique. The PROAFTN method is very efficient for classification and detection of network based intrusion detection system. PROAFTN has several advantages. For example, it uses the multi-criteria decision making paradigm and therefore can be used to gain more understanding about the problem domain. Furthermore, it has direct techniques that can enable a decision maker to adjust its parameters. PROAFTN is also a transparent classification method, that is, its fuzzy component enables the user to have access to more detailed information concerning the classification decision [1]. Additionally, PROAFTN avoids distance measures such as Euclidean distance by com-paring alternatives through scores of different attributes. More-over, it overcomes some difficulties encountered when data has different units or scales and therefore there is no need for data normalization.

III. PARTICLE OF SWARM OPTIMIZATION

In Particle Swarm Optimization [10] optimizes an objective function by undertaking a population based search. The population comprise of possible solutions, named particles, which are metaphor of birds in flocks. These particles are at random initialized and freely fly across the multi-dimensional seek space. During flight, each particle updates its own velocity and position based on the best experience of its own and the entire population. The different steps involved in Particle Swarm Optimization Algorithm are as follows:

Step 1: All particles' velocity and position are randomly place to within pre-defined ranges.

Step 2: Velocity update – At every iteration, the velocities of all particles are updated based on below expression

$$v_i = v_i + c_1 R_1 (p_{i,best} - p_i) + c_2 R_2 (g_{i,best} - p_i) \quad ....(1)$$

where p_i is the position and v_i are the velocity of particle i, $p_{i,best}$ and $g_{i,best}$ is the position with the 'best' objective value found so far by particle i and the entire population respectively; w is a parameter controlling the dynamics of flying; R_1 and R_2 are random variables in the range [0,1]; c_1 and c_2 are factors controlling the related weighting of equivalent terms. The random variables facilitate the PSO with the ability of stochastic searching.

Step 3: Position updating – The positions of all particles are updated according to,

$$p_i = p_i + v_i \quad ...(2)$$

Following updating, p_i should be verified and limited to the allowed range.

Step 4: Memory updating – Update $p_{i,best}$ and $g_{i,best}$ when condition is met,

$$p_{i,best} = p_i \quad if\ f(p_i) > f(p_{i,best})$$

$$g_{i,best} = g_i \quad if\ f(g_i) > f(g_{i,best}) \quad ...(3)$$

Where $f(x)$ is to be optimized and it is a objective function.

Step 5: Stopping Condition–The algorithm repeats steps 2 to 4 until certain stopping circumstances are met, such as a pre-defined number of iterations. Once closed, the algorithm reports the values of g_{best} and $f(g_{best})$ as its solution[8].

PSO utilizes several searching points and the searching points gradually get close to the global optimal point using its pbest and gbest. Primary positions of pbest and gbest are dissimilar However, using thee different direction of pbest and gbest, all agents progressively get close up to the global optimum.

IV. PROPOSED ALGORITHM AND MODEL

In this section discuss the modified algorithm of PROAFTN algorithm with particle of swarm optimization. The particle of swarm optimization used for the selection of feature for the classification of PROAFTN classification method. The process of feature selection define constraints function for multiple set of attribute. The multiple set of attribute of optimal feature selected by particle of swarm optimization. The PROAFTN classification algorithm divide into five level for decision system. These decision system assigned the class of network based intrusion data. The definition of class decide according to the level of attack categories such as DOS, PROB, U2R and R2L. Initially the PROAFTN classification technique decide the uniform categories of data for the classification. The process of classification define in such manner is

```
Initial categories ();
While {
For each dataset set {
For each categories {
For each attribute {
All level = PROAFTN ();
All-level = all-level + level;
}
If (all level > TN) // here TN gives the value of variable value of attribute
Class is attack;
Else
Class is normal;
}
Compare the label class with test class data
}
```


Calculate optimal level for next process of classification
Stored-optimal-feature ()
Selection-process of PROAFTN ()
Voting-process ()
}

1) PROAFTN Algorithm

In our algorithm, we use five level categories for classification of attack in intrusion detection process to form a detection categories. The five level includes five parameters which are a, b, c, d and e. The algorithm calculates the level status of being attacked from the parameters as shown below.

If (a=normal class) && (b=abnormal class) {
Level = dataset – a/ (b-a)
}
Else if (b= normal class and c= abnormal class) {
Level = dataset – b/(c-b)}
Else if (c=normal class and d=abnormal class) {
Level = d – dataset/ (d-c)
Else if (d=normal class and e=abnormal class) {
Level = e– dataset/ (e-d)
Else if (e=normal class and a=abnormal class) {
Level = a– dataset/ (a-e)
}
Else
{
Level =0
}

2) After finding a real level of data we apply POS for the selection of feature for the process of classification.

Input: number of level *X*== *{*a; b,…….,e*}*

Output: set of feature set

1: $G(s)=\frac{N(s)}{D(s)}=\frac{\sum_{i=0}^{n-1}A_i s^i}{\sum_{i=0}^{n}a_i s^i}$ Umpire the termination conditions. If the termination situation are satisfied, then turn to step 9, if not, turn to step 10;
2: Crack to find and compute the optimal feature of data.
3: find final population of POS
4: Take the feature for optimization on population P(i) and generate the next generation A(i+1) . Then turn to step
5: for h ∈ A(i+1) do
6: h.nn ← Nearest-neighbor (A(i+1)- {h})
7: h.sc ← Compute-SC (h, h.nn)
8: V←V ∪{h}
9: V←V ∪{h.nn}
10: if h.sc <th_{sc} then
11: E←E ∪ {(h,h.nn)}
12: endif
13: end for
14: count ← Con-Components (G)

for each pair of components (g1,g2) ∈ G do

15: μ_1←mean-dist (g1), μ_2←mean-dist (g2)
16: if $\frac{\mu_1+\mu_2}{2*\text{centroid\_dist (g1,g2)}}$> 1 then g1←Merge (g1, g2)
17: end for

// Now assign the class labels

18: N_type ← empty
19: for x∈Nlist do
20: h←PseudopointOf(x)//find the corresponding pseudopoint
21: N_type ←N_type∪{(x, h.componentno)}
22: end for
23: data are classified
24: Measure voting process true positive and true negative.

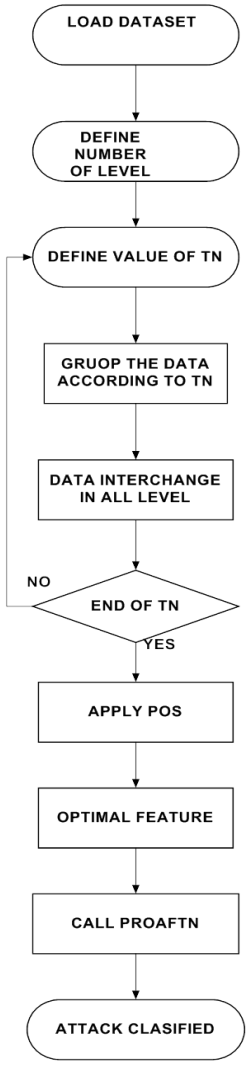

Figure 1 shows that proposed model of PROAFTN classification with POS

V. EXPERIMENTAL RESULT ANALYSIS

In this paper we perform experimental process of proposed classification algorithm for intrusion detection system. The proposed method implements in mat lab 7.14.0 and tested with very reputed data set from UCI machine learning research center. In the research work, I have measured detection accuracy, Precision, recall and F-measure for the SVM, PROAFTN and Proposed method. To evaluate these performance parameters I have used KDDCUP99 datasets from UCI machine learning repository [41] namely intrusion detection dataset.

PERFORMANCE PARAMETERS

Earlier application of isolated feature reduction on dataset has much greater Accuracy, than later by integrating both feature reduction and Improved ID3 Methods. Also there is a considerable enhancement in the true positive and true negative detection ratio and minimizes in false positive and false negative ratio .Thus this gives the direct improvised accuracy in the result. Basis the result of confusion matrix (true positive, true negative, false positive, false negative).We are showing the consequence for the following parameters i.e. - Accuracy, Precision, Recall for data sets.

Precision- Precision measures the proportion of predicted positives/negatives which are actually positive/negative.

Recall -It is the proportion of actual positives/negatives which are predicted positive/negative.

Accuracy-It is the proportion of the total number of prediction that were correct or it is the percentage of correctly classified instances.

Below we are showing how to calculate these parameters by the suitable formulas. And also, below we are showing the graph for that particular data set.

Precision = $\frac{TP}{TP+FP}$

Recall = $\frac{TP}{TP+FN}$

Accuracy = $\frac{TP+TN}{TP+TN+FN+FP}$

Table 1: Shows that the performance evaluation of given input value such as 0.1, 0.5, 0.8 and 0.9 for the classification method such as SVM, PROAFTN and Proposed method.

Input value	Method	Accuracy	Precision	Recall	F-measure
0.1	SVM	89.79	86.27	83.81	85.81
	PROAFTN	95.30	88.70	84.85	86.85
	PROPOSED	96.30	89.80	88.06	88.94
0.5	SVM	91.55	88.03	85.57	87.57
	PROAFTN	97.06	90.45	86.61	88.64
	PROPOSED	98.06	91.59	89.82	89.62
0.8	SVM	93.26	87.29	88.23	89.67
	PROAFTN	95.24	89.76	89.62	90.47
	PROPOSED	96.78	91.56	91.36	91.58
0.9	SVM	94.87	90.48	91.28	91.46
	PROAFTN	95.68	92.57	92.36	92.46
	PROPOSED	98.79	94.78	94.56	94.26

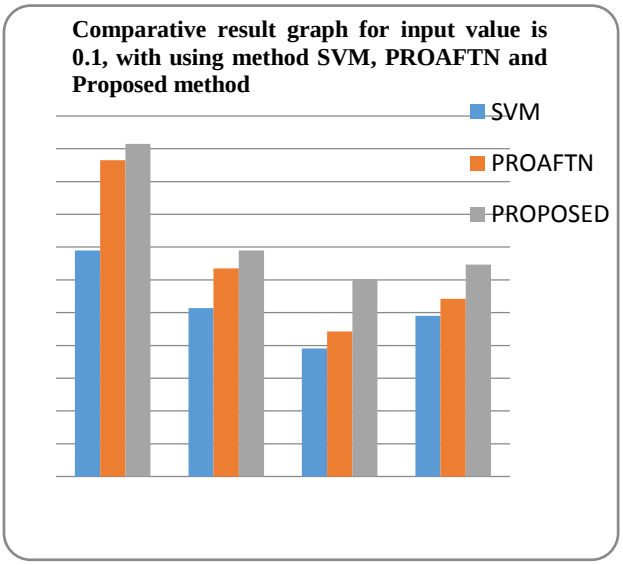

Figure 2: Shows that the comparative result graph for the SVM, PROAFTN and Proposed method and find the Classification Accuracy, Precision, Recall and F-Measure for the given number of input value, and the number of given input value is here 0.1.

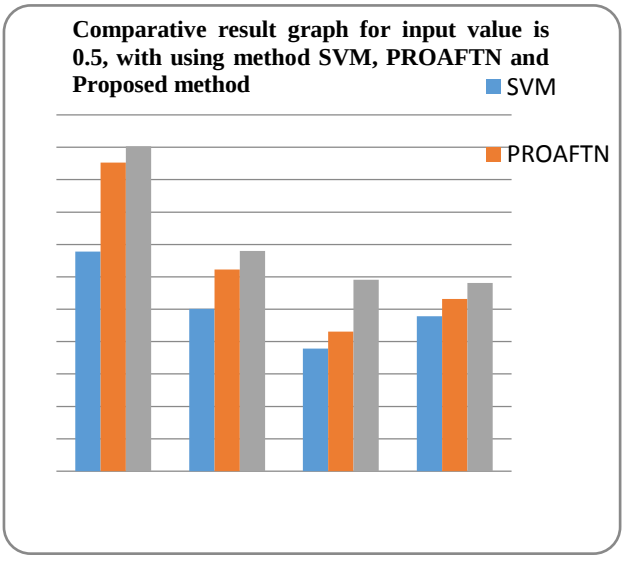

Figure 5.6.2: Shows that the comparative result graph for the SVM, PROAFTN and Proposed method and find the Classification Accuracy, Precision, Recall and F-Measure for the given number of input value, and the number of given input value is here 0.5.

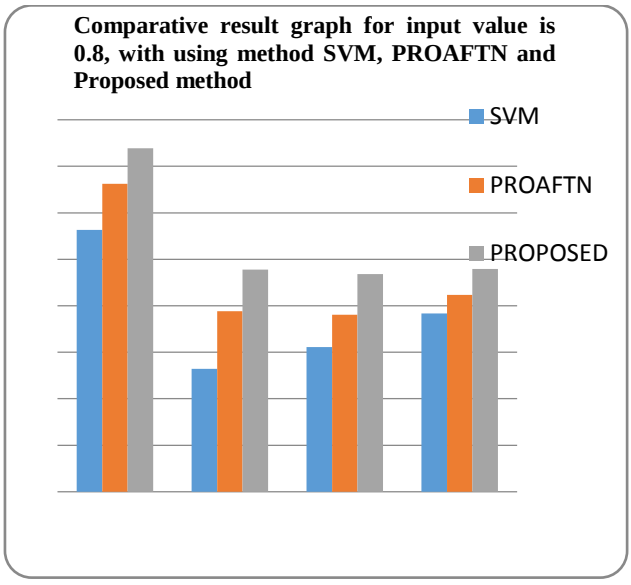

Figure 5.6.3: Shows that the comparative result graph for the SVM, PROAFTN and Proposed method and find the Classification Accuracy, Precision, Recall and F-Measure for the given number of input value, and the number of given input value is here 0.8.

VI. CONCLUSION AND FUTURE WORK

In this paper modified the PROAFTN classification algorithm along with particle of swarm optimization technique. The PROAFTN classification technique directly process data for the classification process, now in this process used feature selection of network file are used. For the selection of feature used particle of swarm optimization technique. Particle of swarm optimization technique gives the optimal feature selection process for the classification of PROAFTN classification method. The PROAFTN classification technique is a hybrid composition of fuzzy logic and protein classification process. The PROAFTN classification gives the multi-criteria support system for network based intrusion detection system. The multi-criteria process of algorithm justify the attack level of different categories of attack. The level of attack identified the valid class for the classification process. The complexity of algorithm also increase now in traffic of network take more time for the filtration of packet. In future also reduces the computational time of modified algorithm.

REFERENCES

[1] Feras N. Al-Obeidat, El-Sayed M. El-Alfy "Network Intrusion Detection Using Multi-Criteria PROAFTN Classification" IEEE, 2014. Pp 1-5.

[2] ShafighParsazad, EhsanSaboori, Amin Allahyar "Fast Feature Reduction in Intrusion Detection Datasets" MIPRO 2012, Pp 1023-1029.

[3] AbebeTesfahun, D. LalithaBhaskari "Intrusion Detection using Random Forests Classifier with SMOTE and Feature Reduction" International Conference on Cloud & Ubiquitous Computing & Emerging Technologies, 2013. Pp 127-132.

[4] HachmiFatma, Limam Mohamed "A two-stage technique to improve intrusion detection systems based on data mining algorithms" IEEE, 2013. Pp 1-6.

[5] Shailendra Singh, Sanjay Silakari "An Ensemble Approach for Cyber Attack Detection System: A Generic Framework" 14th ACIS, IEEE 2013.

[6] Li, "Using Genetic Algorithm for Network Intrusion Detection" Proc. the United States Department of Energy Cyber Security Group 2004 Training Conference, May 2004.

[7] Dewan Md. Farid, Jerome Darmont, NouriaHarbi, Nguyen HuuHoa, Mohammad Zahidur Rahman "Adaptive Network Intrusion Detection Learning: Attribute Selection and Classification" 2008. Pp 1-5.

[8] Gary Stein, Bing Chen, Annie S. Wu, Kien A. Hua "Decision Tree Classifier For Network Intrusion Detection With GA-based Feature Selection" 2556. Pp 1-6.

[9] RituRanjani Singh, Prof. Neetesh Gupta "To Reduce the False Alarm in Intrusion Detection System using self Organizing Map" in International journal of Computer Science and its Applications.

[10] Z. Xue-qin, G. Chun-hua, L. Jia-jin "Intrusion detection system based on feature selection and support vector machine" Proc. First International Conference on Communications and Networking in China (ChinaCom '06), Oct. 2006.

[11] Zhang , M. Zulkernine "Network Intrusion Detection using Random Forests" School of Computing Queen's University, Kingston Ontario, 2006.

[12] John Zhong Lei and Ali Ghorbani "Network Intrusion Detection Using an Improved Competitive Learning Neural Network" in Proceedings of the Second Annual Conference on Communication Networks and Services Research IEEE.

[13] P. Jongsuebsuk,N. Wattanapongsakorn and C. Charnsripinyo "Network Intrusion Detection with Fuzzy Genetic Algorithm for Unknown Attacks" in IEEE 2013.
[14] Deepak Rathoreand Anurag Jain "a novel method for intrusion detection based on ecc and radial bias feed forword network" in Int. J. of Engg. Sci. & Mgmt. (IJESM), Vol. 2, Issue 3: July-Sep.: 2012.
[15] Wing w. Y. Ng, rocky k. C. Chang and daniel s. Yeung "dimensionality reduction for denial of service detection problems using rbfnn output sensitivity" in Proceedings of the Second International Conference on Machine Learning and Cybernetics, Wan, 2-5 November 2003.
[16] AnshulChaturvedi and Prof. VineetRichharia "A Novel Method for Intrusion Detection Based on SARSA and Radial Bias Feed Forward Network (RBFFN)" in international journal of computers & technology vol 7, no 3.
[17] Jain , Upendra "An Efficient intrusion detection based on Decision Tree Classifier using feature Reduction", International Journal of scientific and research Publications , Vol. 2, Jan. 2012.
[18] E. Blanzieri and A. Bryl "Asurvey of learning-based techniques of email spam filtering" Artif. Intell. Rev., vol. 29, no. 1, pp. 63–92, 2008.
[19] D. Sculley and G. Cormack "Filtering email spam in the presence of noisy user feedback" in Proc. 5th Email Anti-Spam Conf., 2008, pp. 1– 10.
[20] Hengjie Li, Jiankun Wang " Intrusion Detection System by Integrating PCNN and Online Robust SVM" IFIP International Conference on Network and Parallel Computing, 2007.pp 250-255.
[21] V. Engen, J. Vincent, and K. Phalp "Enhancing network based intrusion detection for imbalanced data" Int. J. Knowl.-Based Intell. Eng. Syst., vol. 12, no. 5–6, pp. 357–367, 2008.
[22] S. T. Powers and J. He"A hybrid artificial immune system and self organizing map for network intrusion detection" Inf. Sci., vol. 178, no. 15, pp. 3024–3042, 2008.
[23] K. Shafi, T. Kovacs, H. A. Abbass, andW. Zhu "Intrusion detection with evolutionary learning classifier systems" Nat. Comput., vol. 8, no. 1, pp. 3–27, 2009.
[24] Y.Yang and S. A. Elfayoumy "Anti-spam filtering using neural networks and Baysian classifiers" in Proc. IEEE Int. Symp. Comput. Intell. Robot. Autom., 2007, pp. 272–278.
[25] Mohammad Behdad, Luigi Barone, Mohammed Bennamoun and Tim French "Nature-Inspired Techniques in the Context of Fraud Detection" in ieee transactions on systems, man, and cybernetics part c: applications and reviews, vol. 42, no. 6, november 2012.
[26] Alberto Fernandez, Maria Jose del Jesus and Francisco Herrera "On the influence of an adaptive inference system in fuzzy rule based classification system for imbalanced data-sets" in Elsevier Ltd. All rights reserved 2009.
[27] Zonghuazhang, Hong Shen "Application of online-training SVMs for real time intrusion detection with different considerations" Computer Communications, Vol.28, 2005, pp.1428-1442.
[28] WunHwa Chen, ShengHsun Hsu, and HwangPin Shen "Application of SVM and ANN for intrusion detection" Computers & Operations Research, Vol.32, 2005, pp. 2617–2634.
[29] M. Wu, R. C. Miller, and S. L. Garfinkel "Do security toolbars actually prevent phishing attacks" in Proc. SIGCHI Conf. Human Factors Comput. Syst., 2006, pp. 601–610.
[30] S. X. Wu and W. Banzhaf "The use of computational intelligence in intrusion detection systems: A review" Appl. Soft Comput., vol. 10, no. 1, pp. 1–35, 2010.
[31] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez and E.Vazquez "Anomaly-based network intrusion detection: Techniques, Systems and challenges" in Elsevier Ltd. All rights reserved 2008.
[32] Terrence P. Fries "A Fuzzy-Genetic Approach to Network Intrusion Detection" in GECCO 08, July12–16, 2008, Atlanta, Georgia, USA.
[33] ZoranaBankovic, DusanStepanovic,SlobodanBojanic and Octavio Nieto-Taladriz "Improving network security using genetic algorithm approach" in Published by Elsevier Ltd 20367.
[34] Mrutyunjaya Panda and ManasRanjanPatra "network intrusion detection using naive bayes" in IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007.
[35] AnimeshPatcha and Jung-Min Park "An Overview of Anomaly Detection Techniques: Existing Solutions and Latest Technological Trends" in Computer networks 2007.
[36] Ren Hui Gong, Mohammad Zulkernine and PurangAbolmaesumi "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection" in I38EE 2005.
[37] Jonatan Gomez and DipankarDasgupta "Evolving Fuzzy Classifiers for Intrusion Detection" in IEEE 2002.
[38] Francisco Herrera "Genetic fuzzy systems: taxonomy, current research trends and prospects" in Springer-Verlag 2008.
[39] Adel NadjaranToosi and Mohsen Kahani "A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers" in Elsevier B.V. All rights reserved 2007.
[40] C. Phua, V. C. S. Lee, K. Smith-Miles, and R.W. Gayler "A comprehensive survey of data mining-based fraud detection research" CoRR,", pp. 1–14, 2010.
[41] www.uci.com
[42] www.matlab.com

