Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 48 (2015) 338 – 346

Procedia Computer Science

International Conference on Intelligent Computing, Communication & Convergence

(ICCC-2015)

Conference Organized by Interscience Institute of Management and Technology,

Bhubaneswar, Odisha, India

Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection Approach

JABEZ J[a], Dr.B.MUTHUKUMAR[b]*

[a]*Sathyabama Unicersity,Sholinganallur,Chennai*
[b]*Sathyabama Unicersity,Sholinganallur,Chennai*

Abstract

An Intrusion Detection System (IDS) is a software application or device that monitors the system or activities of network for policy violations or malicious activities and generates reports to the management system. A number of systems may try to prevent an intrusion attempt but this is neither required nor expected of a monitoring system. The main focus of Intrusion detection and prevention systems (IDPs) is to identify the possible incidents, logging information about them and in report attempts. In addition, organizations use IDPS for other purposes, like identifying problems with security policies, deterring individuals and documenting existing threats from infringing security policies. IDPS have become an essential addition to the security infrastructure of nearly every organization. Various methods can be used to detect intrusions but each one is specific to a specific method. The main goal of an intrusion detection system is to detect the attacks efficiently. Furthermore, it is equally important to detect attacks at a beginning stage in order to reduce their impacts. This research work proposed a new approach called outlier detection where, the anomaly dataset is measured by the Neighborhood Outlier Factor (NOF). Here, trained model consists of big datasets with distributed storage environment for improving the performance of Intrusion Detection system. The experimental results proved that the proposed approach identifies the anomalies very effectively than any other approaches.

*Keywords:*IDS, Oulier, Anomaly

Peer-review under responsibility of scientific committee of International Conference on Computer, Communication and Convergence (ICCC 2015)
doi:10.1016/j.procs.2015.04.191

1. Introduction

With the high usage of Internet in our day today life, security of network has become the key foundation to all web applications, like online auctions, online retail sales, etc. Detection of Intrusion, attempts to detect the attacks of computer by examining different information records observed in network processes [2] [9]. This can be considered as one of the significant ways to effectively deal with the problems in network security.

An intrusion in the internet can compromise the data security through several internet means. Nowadays, the fast rising networks proliferation, data transfer rate, and an unpredictable Internet usage have added more anomaly problems. Thus researchers need to develop more reliable, effective, and self-monitoring systems, which sort troubles and can carry out operation devoid of human interaction. By undergoing this kind of attempts, catastrophic failures of susceptible systems can be reduced.

Detection stability and detection precision are two key indicators used to evaluate IDS (Intrusion Detection System) [26]. Many of the IDS research studies have been done in order to improve the detection stability and detection precision [22]. In the beginning stage, the research work focus lies in using statistical approaches and rule-based expert systems [17]. But, the results of statistical approaches and rule-based expert systems were not accurate, when encountering larger datasets. In order to overcome the abovementioned problem, many data mining techniques were developed [7] .

Some machine-learning paradigms containing Linear Genetic Programming (LGP) [19], neural networks [18], Bayesian networks, Support Vector Machines (SVM), Fuzzy Inference Systems (FISs) [25], Multivariate Adaptive Regression Splines (MARS) [20] etc., have been investigated for the design of Intrusion Detection System (IDS). Thus, one of the most common techniques in machine-learning paradigms is known as Neural Network (NN) that should be used for resolving a lot of complex practical problems which has been successfully applied into Intrusion Detection System [9]. Nevertheless, the major drawbacks of Neural Network-based IDS exist in two features:

1. Lower Detection Precision- particularly for low-frequent attacks, e.g., U2R (User to Root), R2L (Remote to Local).
2. Weaker detection stability [4].

To solve the above two problems, this research work propose a novel approach for outlier computation-based IDS, Outlier Detection Approach, to enhance the detection precision for low-frequent attacks and detection stability. The proposed approach has got two stages such as training with normal big datasets and testing with intrusion datasets. A set of various big datasets are used to train our IDS in the initial stage at distributed storage environment. Normal big datasets are improving the performance of Intrusion Detection System. Assume an intrusion dataset which is used to compute an error value with trained big data sets. If number of error value is increased such as the specified threshold then the tested data set consider as anomaly dataset.

The rest of the paper is organized as follows: Section II explains the existing work. Next, Section III provides the details of the concept and classification of normal intrusion detection system components and its proposed approach. Section IV shows the proposed approach, the experimental results and its analysis. Finally, Section V concludes the work and its future directions.

2. Literature Survey

This section deals with the attempts made by researcher in the area of network based intrusion detection system and most of the detection works were based on KDD dataset. An expert system based on rules and statistical approaches are the two commonly used approaches to ensure intrusion detection. The Expert system based on rules will detect the known intrusion in high rate and it will not identify new intrusion. Where, the database should be continuously updated. In statistical approach, Intrusion Detection System includes different methods like Cluster analysis, Multivariate analysis, Bayesian analysis, and Principal component analysis. Many new techniques from data mining should be proposed to overcome the problems of above mentioned approaches. Many results are produced in the KDD cup 99 dataset research work and they are briefly discussed.

Anderson [25] suggested an intrusion detection method to efficiently detect the intrusion. An Intrusion Detection Mechanism using Time- series, Markov chains, and statistics was developed by Denning [3] Denning considered that the changes in the normal behavior of user are treated as anomalous. For monitoring and detecting

user's events an Expert System of intrusion detection was developed by Stanford Research Centre. This centre also developed next generation mechanism which includes audit profiles of user's and can monitor the current status of the user, if any change occurs with user's activity compared with audit profile of user then it will generate am alarm. Haystack [22] later developed a framework to estimate an intrusion detection method based on user and anomaly strategies. Six types of intrusion were detected and those includes the masquerade attacks, malicious use, leakage, service denial, unauthorized user's break- ins attempt, and access control of security system. The source fire developed indicates a network based intrusion detection and prevention mechanism called SNORT system which is an open source. Forrest [10] in 1996 created a normal profile based on analyzing the call sequences between intrusion detection and protection against human system. An attack in this system is considered as the sequence deviation from normal profile sequence. Thus, this system works offline using previously collected information and implements view table algorithm for learning program profiles significantly.

Duan et al. [8] have concentrated on identifying compromised machines that are recruited to detect spam zombies. An approach SPOT is proposed to scan sequentially outgoing messages by implementing SPRT (Sequential Probability Ratio Test). This method quickly estimates whether a host is compromised or not. Identifying compromised machines using malware infection system is stated by Bot hunter [13]. This system has large no of steps that allow intrusion detection alarms correlation triggered using inbound traffic with outgoing message exchange pattern results. Bot Sniffer [14] explained in his work about compromised machine characteristics which are a uniform temporal-spatial behavior for detecting zombies. This method identifies zombies by combining flows based on server connections and searching flows with similar behavior respectively.

Kumar and Goyal [12] have explained implements genetic algorithms in dataset training to classify the labels that are smurf attacked and achieves low false positive ratio of 0.2%. Further work done by Abdullah [1] and co-workers elaborated intrusion detection classification rules using genetic algorithms. Intrusion detection rules using genetic algorithms was also the study made by Ojugo et al. [21]. This method uses fitness function for estimating the rules.

Machine learning techniques are also implemented to detect the intrusion. Existing machine learning techniques (Artificial Neural Networks - ANN) for intrusion detection was described by Roshani team [23].

Gaikwad et al [11] introduced a technique based on fuzzy clustering and ANN approach. This method could be applicable to overcome the issues of weak stability detection as well as low precision detection. The restore point in this method was employed for registry keys, system files roll back, project database and installed programs. Fuzzy clustering will generate different subsets for training in order to reduce the amount of subset size and complexity. Then each subset is trained with different type of artificial neural network and finally processed to obtain significant results. Jaiganesh et al [15] suggested a novel back propagation model for intrusion detection. This method makes training pair with a combination of input and equivalent target were generated and implemented into the network. Performance success can be measured by false alarm and detection rate. Detection rate was proven to be less than 80% for U2R, R2L, DoS and Probe attacks. However, the major issue of the method was found to be much inefficient to detect hidden attackers present in the system. Devikrishna et al [5] used MLP (Multi Layer Perceptron) architecture for intrusion detection that detects and classifies attacks into six types. MLP method was considered as a failure model due to irrelevant output .In the present paper we have tried to overcome this query and to establish a better detection technique .

Lin GU et al [16] proposed empirical study for right choice of unstable growing demand in processing big data which entails huge burden of storage, data center communication and computation which brings substantial operational expenditure for data providing centers. Apart from traditional cloud service, an important characteristic of big data was found to be the tight coupling of computation and data computation tasks were performed only with relevant data. But the means to improve the IDS is not clearly conveyed so far by any of the researchers. Thus, the main aim of this paper is to implement a clear picture of the IDS using distributed big data concept.

Issues of existing techniques

Many issues are been stated in the existing literature survival like additional training time, accurate identification of low common attacks and attacks classification. In order to solve the issue of additional training time, it is must to develop a new high-speed algorithm for intrusion detection system and its results will be tested with existing techniques. In contrast to the existing approaches that performed some kind of inefficiency in intrusion detection, the main aim of our research work is to propose a new high speed algorithm for reducing training time. The obtained results are also to be discussed along with the existing methods.

3. Intrusion Detection System

3.1 Classification of Intrusion Detection Systems (IDS)

Classification is one of the best – known solution approaches. National Institute of Standards and Technology (NIST) organization provides guidance document on Intrusion Detection Systems [24].
Intrusion Detection System briefly classified into three different categories:

- **Host-based IDS**
- **Network-based IDS**
- **Vulnerability-assessment IDS**

There are two basic models used to analyze the events and discover attacks:

- **Misuse detection model** – Intrusion Detection System detect intrusions by looking for similar activities such as vulnerabilities or known intrusion signatures.
- **Anomaly detection model -** IDS detect intrusions by searching « abnormal » network traffic.

The misuse detection model is commonly referred as IDS commercial tool; always Vendors must update intrusion signatures. Anomaly detection based IDS model have the capability to detect attack symptoms without specifying attack models, but these models are very sensitive to false alarms. In the present study we have utilized the proposed IDS approach's based on the anomaly detection model.

3.2 System Architecture

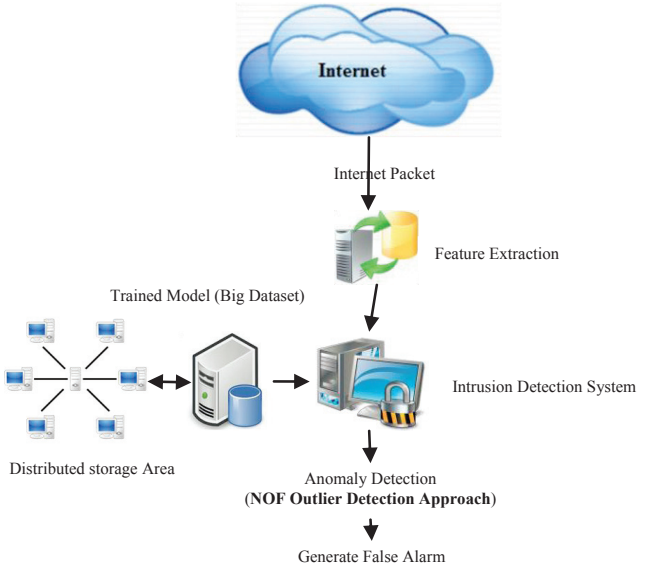


Fig. 1: Proposed System Architecture

Our main aim is to develop an IDS based on anomaly detection model that would be precise, not easily cheated by small variations in patterns, low in false alarms, adaptive and be of real time. The Figure 1 describes the proposed system architecture were the intrusion packets are received from the internet then SNORT is used to collect the datasets. Initially, the features extracted from data packets then it forwarded to our proposed IDS. Then, proposed IDS compute the distance between the extracted features and trained model. Here, trained model consists of big datasets with distributed storage environment to improve the performance of Intrusion Detection system. Thus, the outlier value is greater than the specified threshold then it generates the false alarm.

3.2 Pseudocode: Outlier Detection Approach

The normal data objects have a dense neighborhood whereas "outliers" are far apart from their neighbors. The outliers are the objects of the outer layers. The major idea behind this approach is to assign a data example to being outlier degree called Neighborhood Outlier Factor (NOF) and to find the rare data whose behavior is very exceptional when compared with large amount of normal data. The algorithm steps used to calculate NOFs for all data examples are as follows:

1. For each data example O, calculate the k-distance is the nearest neighborhood (where all points in a k-distance forms sphere).

2. Next, calculate the reachability distance for every data example O with respect to data example p as: reach-distance (O,p) = max{k-distance (p), d(O,p)}, where d(O,p) is the distance between data example O and data example p.

3.Then, calculate local reachability density for each data example O , inverse of the average reachability distance is based on the MinPts (minimum number of objects) data example O and its nearest neighbors.

4. Calculate NOF to all data example O as an average of the data example O's local reachability density ratios and local reachability density of O's MinPts nearest neighbors.

The benefits of proposed NOF approach is illustrated in Figure 2. Clusters are formally defined as maximal sets of density-connected objects. Here a simple two-dimensional dataset is taken with much larger number of examples in cluster C1 then C2. So the cluster density of C2 is extensively higher than that of C1 cluster density. For each example consider an object q inside the cluster C1, the distance between the example q and its nearest neighbor is greater than the distance between the example p2 and the nearest neighbor from the cluster C2, and the example p2 will not be considered as outlier.

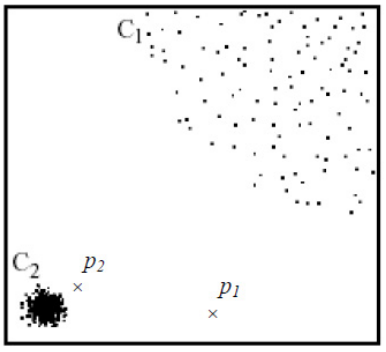


Fig. 2: NOF Outlier Detection Approach

Consequently, the outlier detection lies in the field of statistics. Nevertheless, the example p1 can be detected as outlier using only the nearest neighbor distances. Alternatively, NOF is capable to capture both (p1 and p2) outliers due to the fact that it considers the density all round the points.

4. Results and Discussion

4.1 Experimental Setup

In our study, a dataset is extracted and number of experiments were based on the extracted dataset in order to measure the IDS performance. Experiments were carried out based on the following configuration: Windows 7, Intel Pentium (R), CPU G2020 and processer speed 2.90 GHz respectively.

The extracted data set includes training data of about two thousand connection records and test data includes five thousand connection records. In addition, dataset includes a group of forty one derived features

received from every connection and also a group of labels that identifies the connection record status whether it is a normal type or attacked type.

4.2 Anomaly Results

Table 1: Fragmentation of Attributes from the IP datasets

S. No	Attributes	S.No	Attributes
1	Duration	6	Destination bytes
2	Protocol type	7	Number failed logins
3	Service	8	Service received error rate
4	Flag	9	Different service rate
5	Source bytes	10	Destination host count

List of partial attributes names obtained from network datasets are as shown in Table 1. These attributes detects whether the received network dataset is anomaly or not.

Table 2: Fragmentation of trained normal big-data set model

ID	Duration	Flag	Source byte	Destination byte		ID	Duration	Flag	Source byte	Destination byte
1.	81	18	522	0		6.	66	28	522	0
2.	12	61	0	0		7.	78	132	18	0
3.	22	61	0	0		8.	45	134	0	0
4.	65	184	520	0		9.	74	58	89	0
5.	45	47	0	0		10.	35	1	50	0

Dataset's are obtained from different communication level network with different internet service provider's policy. Internet service provider policy will vary for different communication levels (Table 2).

Table 3: Fragmentation of information Received from various user

ID	Duration	Flag	Source byte	Destination byte	ID	Duration	Flag	Source byte	Destination byte
1.	10	SF	491	0	6.	98	233	616	0
2.	22	334	0	0	7.	569	147	105	0
3.	56	146	0	0	8.	45	RSTR	0	0
4.	78	199	420	0	9.	87	255	861	0
5.	66	28	0	0	10.	35	1	0	0

The network information of user's will vary because different user's use different internet service providers. (Table 3)

Table 4: Distance and Outlier value of tested data

ID	Distance	Outlier value	ID	Distance	Outlier value
1.	2.5	5	6.	1.2	2
2.	4.6	8	7.	2.7	3
3.	3.6	7	8.	4.2	4
4.	5.6	10	9.	2.7	3
5.	2.4	4	10.	2.9	5

The distance and outlier values of tested data which is calculated by proposed outlier detection method. It indicates that outlier values increase if distance between the normal and tested dataset increases. The results are shown in Table 4.

4.3 Discussion

4.3.1 Comparison of proposed approach and Existing approach (Execution Time Vs Dataset Size)

Fig 3: Big-Dataset size Vs Execution Time

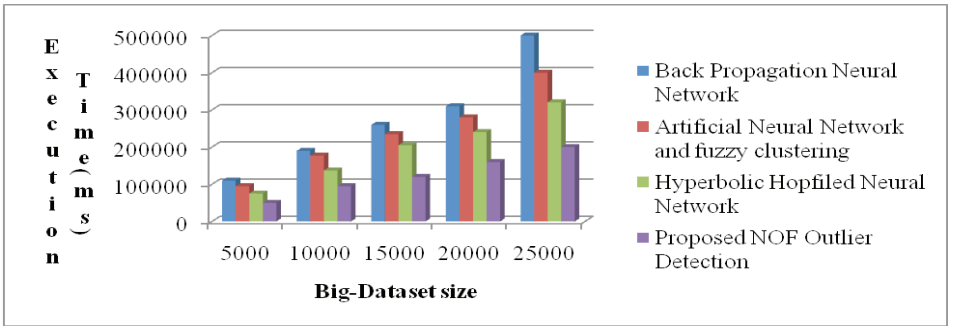


Figure 3 shows the overview of various execution times with various size of dataset. The proposed Intrusion Detection System takes less execution time at every level rather than other existing machine learning approaches. The cause is less trained datasets thus the distance computation is easy between the trained and testing dataset respectively

4.3.2 Comparison of proposed approach and Existing approach (Anomaly Detection Rate Vs Dataset Size)

Fig 4: Big-Dataset size Vs Anomaly Detection

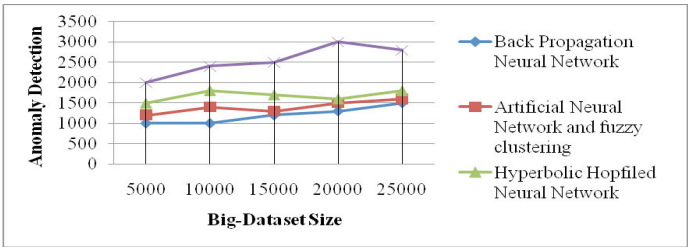
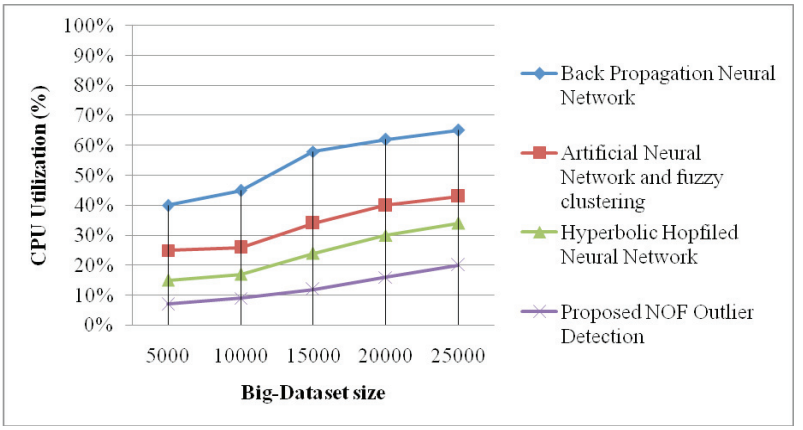


Fig.4 shows the anomaly detection rate in the computer network. The proposed Intrusion Detection System identifies almost all type of attacks such as Probe, DoS, U2R and R2L. The anomaly detection rate depends on the outlier values testing data. If the outlier value increase then the dataset assumed acts as intrusion dataset.

4.3.3 Comparison of proposed approach and Existing approach (CPU Utilization Vs Dataset Size)

Fig 5: Big-Dataset size Vs CPU Utilization



The figure 5 shows the graphical comparison of CPU utilization levels with various sizes of datasets. In the machine learning approaches', CPU utilization is very high when compared with proposed approach. Most of the research papers have assigned machine learning approaches only with the help of huge quantity of training datasets and training functions. In our proposed approach we are using only limited datasets to train the proposed IDS.

5. Conclusion

In this paper, we have presented the details of a new approach called Outlier Detection approach to detect the intrusion in the computer network. Our training model consists of big datasets with distributed environment that improves the performance of Intrusion detection system. The proposed approach is also been tested with the KDD datasets that are received from real world. The machine learning approaches detect the intrusion in the computer network with huge execution time and storage to predict the when compared to the proposed IDS system which takes less execution time and storage to test the dataset .Here in this study, the performance of proposed IDS is better than that of other existing machine learning approaches and can significantly detect almost all anomaly data in the computer network. In future, the proposed work can be possibly used for various distance computation function between the trained model and testing data. Our research work can be considered to improve the efficiency of IDS in a better manner.

References:

1. Abdullah, B., Abd-algafar I., Salama G. I. and Abd-alhafez A. Performance Evaluation of a Genetic Algorithm Based Approach to Network Intrusion Detection System, Proceedings of 13th International Conference on Aerospace Sciences and Aviation Technology (ASAT-13), Military Technical College, Cairo, Egypt, 2009;1-5.
2. Anderson, J. P. Computer security threat monitoring and surveillance. Technical Report, Fort Washington, PA, USA.,1980;9-11.
3. Anderson, D., Frivold, T. and Valdes, A. Next-generation intrusion detection expert system (NIDES): A summary Technical Report SRI–CSL–95–07,Computer Science Laboratory,SRI International, May 1995.
4. Beghdad, R. Critical study of neural networks in detecting intrusions. Computers and Security, 27(5-6): 2008;168–175.
5. Devikrishna, K. S. and Ramakrishna , B. B. .An Artificial Neural Network based Intrusion Detection System and Classification of Attacks", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, Jul-Aug 2013, 3(4): 1959-1964.
6. Denning, D. E.. An intrusion detection model, IEEE Transactions on Software Engineering, CA,. IEEE Computer Society Press;1987.

7. Dokas, P., Ertoz, L., Lazarevic, A., Srivastava, J.and Tan, P. N. Data mining for network intrusion detection. Proceeding of NGDM, 2002;21–30.
8. Duan, Z., Chen, P., Sanchez, F., Dong, Y., Stephenson, M. and J. M. Barker, M. (2012). Detecting spam zombies by monitoring outgoing messages, IEEE Trans. Dependable and Secure Computing, Apr 2012; 9(2):198–210
9. Endorf, C., Schultz, E. and Mellander, J. (2004). Intrusion detection and prevention. California: Mc Graw-Hill.
10. Forrest, S., Hofmeyr, S. A. , Somayaji, A. and Longstaff, T. A. A Sense of Self for Unix Processes, IEEE Symposium on Research in Security and Privacy, Oakland, CA, USA, 1996;120--128.
11. Gaikwad, Sonali Jagtap, D.P. Kunal Thakare and Vaishali Budhawant. Anomaly Based Intrusion Detection System Using Artificial Neural Network and fuzzy clustering., International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, November- 2012; 1(9).
12. Goyal, A. and Kumar, C. .GA-NIDS: A Genetic Algorithm based Network Intrusion Detection System, Electrical Engineering and Computer Science, North West University, Technical Report;2008.
13. Gu, G., Porras, P., Yegneswaran V., Fong, M. and Lee, W. BotHunter: detecting malware infection through IDS-driven ialog correlation, Proc. of 16th USENIX Security Symp. (SS '07), Aug. 2007; 12:1–12:16.
14. Gu, G., Zhang, J. and Lee, W. (2008). BotSniffer: detecting botnet command and control channels in network traffic, Proc. of 15th Ann. Network and Distributed Sytem Security Symp. (NDSS '08), Feb. 2008.
15. Jaiganesh, V., Sumathi, P. and Mangayarkarasi, S. ,An Analysis of Intrusion Detection System using back propagation neural network, IEEE Computer Society Publication;2013.
16. Lin Gu, Deze Zeng, Peng Li, and Song Guo. Cost Minimization for Big Data Processing in Geo-Distributed Data Centers, IEEE Transactions on Emerging Topics in Computing;2014.
17. Manikopoulos, C.and Papavassiliou, S. Network intrusion and fault detection: A statistical anomaly approach. IEEE Communications Magazine, 40(10);2002 76–82.
18. Mukkamala, S., Sung, A.H., Abraham, A. Intrusion detection using ensemble of soft computing paradigms, third international conference on intelligent systems design and applications, intelligent systems designand applications, advances in soft computing. Germany: Springer,2003; 239–248.
19. Mukkamala, S., Sung, A.H., Abraham, A. Modeling intrusion detection systems using linear genetic programming approach, The 17th international conference on industrial & engineering applications of artificial intelligence and expert systems, innovations in applied artificial intelligence. In: Robert O., Chunsheng Y., Moonis A., editors. Lecture Notes in Computer Science, Germany: Springer; 2004a. 3029: 2004;633–642.
20. Mukkamala, S., Sung, A.H., Abraham, A. and Ramos ,V. Intrusion detection systems using adaptive regression splines. In: Seruca I, Filipe J, Hammoudi S, Cordeiro J, editors. Proceedings of the 6th international conference on enterprise information systems, ICEIS', Portugal. 2004b. 3: 2004;26–33.
21. Ojugo, A. A., Eboka, A. O., Okanta, O. E., Yora, R. E. and Aghware, F. O.Genetic Algorithm Rule-Based Intrusion Detection System (GAIDS), Journal of Emerging Trends in Computing and Information Sciences, 3(8);2012; 1182 – 1194.
22. Patcha, A. and Park, J. M. An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer Networks, 51(12);2007; 3448–3470.
23. Roshani Gaidhane, Vaidya, C. and Raghuwanshi, M. Survey.Learning Techniques for Intrusion Detection System (IDS), International Journal of Advance Foundation and Research in Computer (IJAFRC) Feb 2014. ISSN 2348 – 4853, 2014;1(2).
24. Planquart, J.P. (2001). Article paper ,Application of Neural Networks to Intrusion Detection, SANS Institute .1-3.
25. Shah, K., Dave, N., Chavan, S., Mukherjee, S., Abraham, A. and Sanyal S. Adaptive neuro-fuzzy intrusion detection system. IEEE International Conference on Information Technology: Coding and Computing (ITCC'04), vol. 1. USA: IEEE Computer Society;2004; 70–74.
26. Silva, L. D. S., Santos, A. C., Mancilha, T. D., Silva, J. D. and Montes, A. Detecting attack signatures in the real network traffic with ANNIDA. Expert Systems with Applications, 34(4);2008; 2326–2333.