HYBRID METHODOLOGY FOR ENHANCING THE POTENTIALITY OF IDS

Vanita M Kewlani[1], Jay D. Amin[2]

[1] *PG Student, Information Technology, LJIET, Ahmedabad, Gujarat, India*
[2] *Assistant Professor, Information Technology, LJIET, Ahmedabad, Gujarat, India*

ABSTRACT

Nowadays, with the drastic progress in communications and networks, protection have turn out to be a primary subject for computer programs. An effective way to discover the unlawful customers is to keep monitoring these person's packets. One-of-a-kind algorithms, ways and applications are created and applied to remedy the crisis of detecting the attacks in intrusion detection techniques. In this paper, we reward an intrusion detection model based on genetic algorithm and neural network. The key proposal is to take competencies of classification advantage of genetic algorithm and neural network for intrusion detection process. The new manne quin has capability to identify an attack, to distinguish one attack from other i.E. Classifying attack , and the importantly,low false negative. This strategy uses evolution idea to information evolution as a way to filter the traffic data and consequently cut down the complexity. To put into effect and measure the efficiency of this procedure. We used the KDD99 benchmark dataset and acquired cheap detection price.

Keyword : *- Genetic Algorithm, Intrusion Detection System, KDD Cup 1999 Dataset, Neural Network ;*

1. INTRODUCTION

With the rapid growth of the internet, computer attacks are increasing at a fast pace and can easily cause millions of dollar in damage to an organization. Detection of these attacks is an important issue of Computer security. Intrusion Detection Systems (IDS) technology is an effective approach in dealing with the problems of network security. In general, the techniques for Intrusion Detection (ID) fall into two major categories depending on the modeling methods used: misuse detection and anomaly detection. Misuse detection compares the usage patterns for knowing the techniques of compromising computer security. Although misuse detection is effective against known intrusion types; it cannot detect new attacks that were not predefined. Anomaly detection, on the other hand, approaches the problem by attempting to find deviations from the established patterns of usage. Anomaly detection may be able to detect new attacks. However, it may also cause a significant number of false alarms because the normal behavior varies widely and obtaining complete description of normal behavior is often difficult. Architecturally, an intrusion detection system can be categorized into three types**:** host based IDS, network based IDS and hybrid IDS [1] [2]. A host based intrusion detection system uses the audit trails of the operation system as a primary data source.

A network based intrusion detection system, on the other hand, uses network traffic information as its main data source. Hybrid intrusion detection system uses both methods [3]. However, most available commercial IDS's use only misuse detection because most developed anomaly detector still cannot overcome the limitations (high false positive detection errors, the difficulty of handling gradual misbehavior and expensive Computation [4]). This trend motivates many research efforts to build anomaly detectors for the purpose of ID [5]. The main problem is the difficulty of distinguishing between natural behavior and abnormal behavior in computer networks due to the significant overlap in monitoring data. This detection process generates false alarms resulting from the Intrusion Detection based on the Anomaly Intrusion Detection System. The use of Genetic algorithm might reduce the amount

of false alarm, where Genetic algorithm is used to separate this overlap between normal and abnormal behavior in computer networks.

2. PREVIOUS WORK

In particular several Neural Networks based approaches were employed for Intrusion Detection. Several Genetic Algorithms (GAs) has been used for detecting Intrusions of different kinds in different scenarios [6][7] [8] [9]. GAs used to select required features and to determine the optimal and minimal parameters of some core functions in which different AI methods were used to derive acquisition of rules [10] [11] [12]. In [13], authors presented an implementation of GA based approach to Network Intrusion Detection using GA and showed software implementation. The approach derived a set of classification rules and utilizes a support-confidence framework to judge fitness function. In [14], authors designed a GA based performance evaluation algorithm for network intrusion detection. The approach uses information theory for filtering the traffic data. In [15], authors used the BP network with GAs for enhancement of BP, they used some types of attack with some features of KDD data. A back-propagation Neural Network was used [16], authors used all features of KDD data, the classification rate for experiment result for normal traffic was 100%, known attacks were 80%, and for unknown attacks were 60%.

3. GENETIC ALGORITHM

3.1 Overview

A Genetic Algorithm (GA) is a programming technique that mimics biological evolution as a problem-solving strategy [17]. It is based on Darwinian's principle of evolution and survival of fittest to optimize a population of candidate solutions towards a predefined fitness [7].

GA uses an evolution and natural selection that uses a chromosome-like data structure and evolve the chromosomes using selection, recombination and mutation operators [7]. The process usually begins with randomly generated population of chromosomes, which represent all possible solution of a problem that are considered candidate solutions. From each chromosome different positions are encoded as bits, characters or numbers. These positions could be referred to as genes. An evaluation function is used to calculate the goodness of each chromosome according to the desired solution; this function is known as "Fitness Function". During the process of evaluation "Crossover" is used to simulate natural reproduction and "Mutation" is used to mutation of species [7]. For survival and combination the selection of chromosomes is biased towards the fittest chromosomes.

When we use GA for solving various problems three factors will have vital impact on the effectiveness of the algorithm and also of the applications [18].

i) The fitness function;
ii) The representation of individuals
iii) The GA parameters.

The determination of these factors often depends on applications and/or implementation.

3.2. Flowchart

Fig 1 shows the operations of a general genetic algorithm according to which GA is implemented into our system. Also all the three steps of generating new population from old population are depicted. The process of generating new population from old population includes selection, crossover, and mutation. If new population is not feasible then quit, otherwise again repeat the generation process.

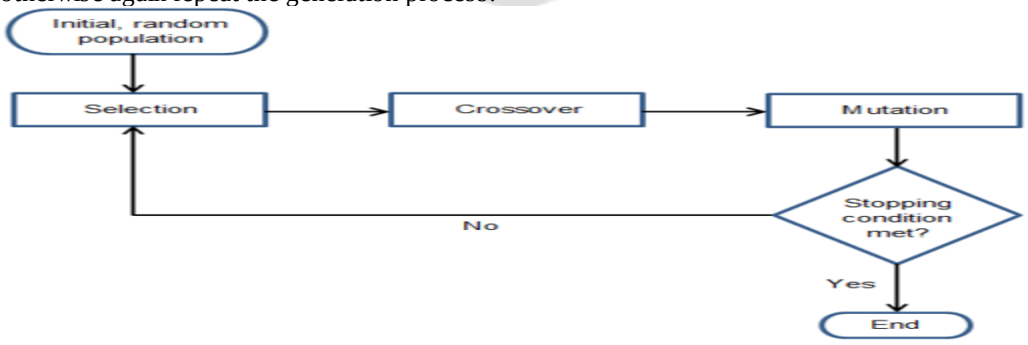

Fig 1:Genetic Algorithm

3.3. Steps of Precalculation System

This system can be divided into two main phases: the precalculation phase and the detection phase. Follo wing are the major steps in precalculation phase, where a set of chromosome is created using training data. This chromosome set will be used in the next phase for the purpose of comparison.

Major Steps in Precalculation
Algorithm: Initialize chromosomes for comparison
Input: Network audit data (for training)
Output: A set of chromosomes
1. Range = 0.125
2. For each training data
3. If it has neighboring chromosome within Range
4. Merge it with the nearest chromosome
5. Else
6. Create new chromosome with it
7. End if
8. End for

4. NEURAL NETWORK

Neural Networks (NNs) have attracted more attention compared to other techniques. That is mainly due to the strong discrimination and generalization abilities of Neural Networks that utilized for classification purposes [19]. Artificial Neural Network is a system simulation of the neurons in the human brain [20]. It is composed of a large number of highly interconnected processing elements (neurons) working with each other to solve specific problems. Each processing element is basically a summing element followed by an active function. The output of each neuron (after applying the weight parameter associated with the connection) is fed as the input to all of the neurons in the next layer. The learning process is essentially an optimization process in which the parameters of the best set of connection coefficients (weights) for solving a problem are found [21].

5. EXPERIMENTAL DESIGN

Genetic Algorithm is a global optimized algorithm with selection in probability as mainly method.Introducing the thinking of GA into neural network can let Back Propogation neural network avoid going to the local minimum. Combining global optimization ability of GA with instructive search ideas of BP algorithm, not only can overcome the blindness of optimization, but avoid occurring local convergence. The application of combining the two methods in the network intrusion detection system will improve the detection efficiency.

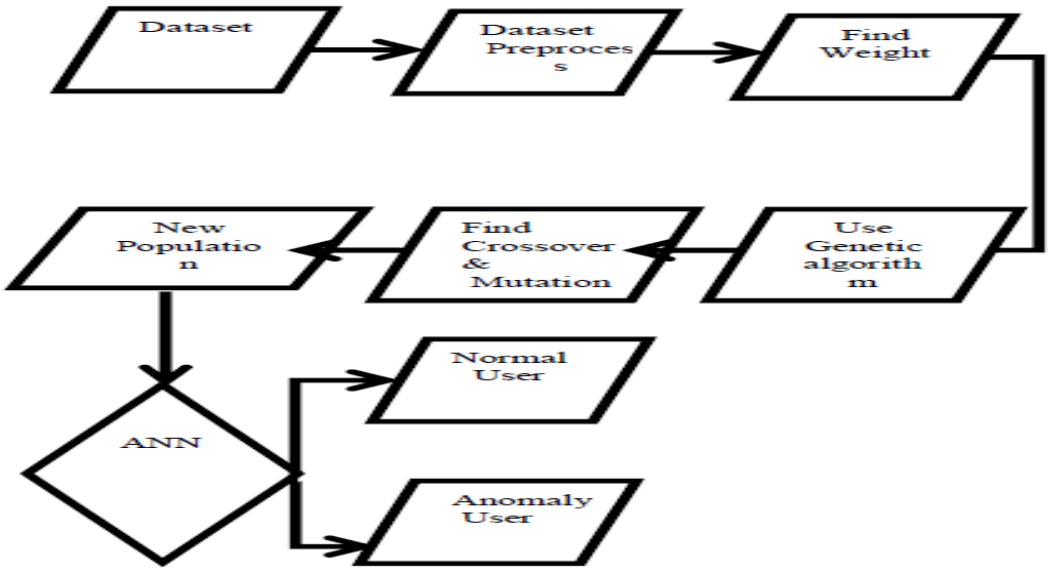

Fig 2:Proposed Flow Diagram

6. KDD DATA SET

KDD 99 data set are used as the input vectors for training and validation of the tested neural network. It was created based on the DARPA intrusion detection evaluation program. MIT Lincoln Lab that participates in this program has set up simulation of typical LAN network in order to acquire raw TCP dump data. They simulated LAN operated as a normal environment, which was infected by various types of attacks. The raw data set was processed into connection records. For each connection, 41 various features were extracted. Each connection was labeled as normal or under specific type of attack.

```
0,tcp,http,SF,181,5450,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,9,9,1.00,0.00,0.11,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,239,486,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,19,19,1.00,0.00,0.05,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,235,1337,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,29,29,1.00,0.00,0.03,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,219,1337,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,6,6,0.00,0.00,0.00,0.00,1.00,0.00,0.00,39,39,1.00,0.00,0.03,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,217,2032,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,6,6,0.00,0.00,0.00,0.00,1.00,0.00,0.00,49,49,1.00,0.00,0.02,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,217,2032,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,6,6,0.00,0.00,0.00,0.00,1.00,0.00,0.00,59,59,1.00,0.00,0.02,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,212,1940,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,2,0.00,0.00,0.00,0.00,1.00,0.00,1.00,1,69,1.00,0.00,1.00,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,159,4087,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,5,5,0.00,0.00,0.00,0.00,1.00,0.00,0.00,11,79,1.00,0.00,0.09,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,210,151,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,8,89,1.00,0.00,0.12,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,212,786,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,8,99,1.00,0.00,0.12,0.05,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,210,624,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,18,18,0.00,0.00,0.00,0.00,1.00,0.00,0.00,18,109,1.00,0.00,0.06,0.05,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,177,1985,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.00,0.00,0.00,1.00,0.00,0.00,28,119,1.00,0.00,0.04,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,222,773,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,11,11,0.00,0.00,0.00,0.00,1.00,0.00,0.00,38,129,1.00,0.00,0.03,0.04,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,256,1169,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,4,4,0.00,0.00,0.00,0.00,1.00,0.00,0.00,4,139,1.00,0.00,0.25,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,241,259,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.00,0.00,0.00,1.00,0.00,0.00,14,149,1.00,0.00,0.07,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,260,1837,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,11,11,0.00,0.00,0.00,0.00,1.00,0.00,0.00,24,159,1.00,0.00,0.04,0.04,0.00,0.00,0.00,0.00,norma
0,tcp,http,SF,241,261,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,2,2,0.00,0.00,0.00,0.00,1.00,0.00,0.00,34,169,1.00,0.00,0.03,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,257,818,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,12,12,0.00,0.00,0.00,0.00,1.00,0.00,0.00,44,179,1.00,0.00,0.02,0.03,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,233,255,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,2,8,0.00,0.00,0.00,0.00,1.00,0.00,0.25,54,189,1.00,0.00,0.02,0.03,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,233,504,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,7,7,0.00,0.00,0.00,0.00,1.00,0.00,0.00,64,199,1.00,0.00,0.02,0.03,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,256,1273,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,17,17,0.00,0.00,0.00,0.00,1.00,0.00,0.00,74,209,1.00,0.00,0.01,0.03,0.00,0.00,0.00,0.00,norma
0,tcp,http,SF,234,255,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,5,5,0.00,0.00,0.00,0.00,1.00,0.00,0.00,84,219,1.00,0.00,0.01,0.03,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,241,259,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,12,12,0.00,0.00,0.00,0.00,1.00,0.00,0.00,94,229,1.00,0.00,0.01,0.03,0.00,0.00,0.00,0.00,normal
0,tcp,http,SF,239,968,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,3,3,0.00,0.00,0.00,0.00,1.00,0.00,0.00,3,239,1.00,0.00,0.33,0.03,0.00,0.00,0.00,0.00,normal.
```

Fig 3:Kddcup Dataset

7. SIMULATION STUDY

We have used JAVA to implement the proposed work also, we have used the base of the standard dataset used in KDD Cup 1999 "Computer network intrusion detection" competition. The Small-KDD is a reduced version of the KDD'99 dataset. The Small-KDD has the same features as the KDD'99 but it does not include the redundant records of the KDD'99, and there are also no duplicate records which make it unbiased to frequent and redundant entries.

Table 1: KDD Dataset:

	NORMAL	ANOMALY	TOTAL
Train("kddcup.data_10_percent")	92	68	160
Test("corrected")	88	53	141

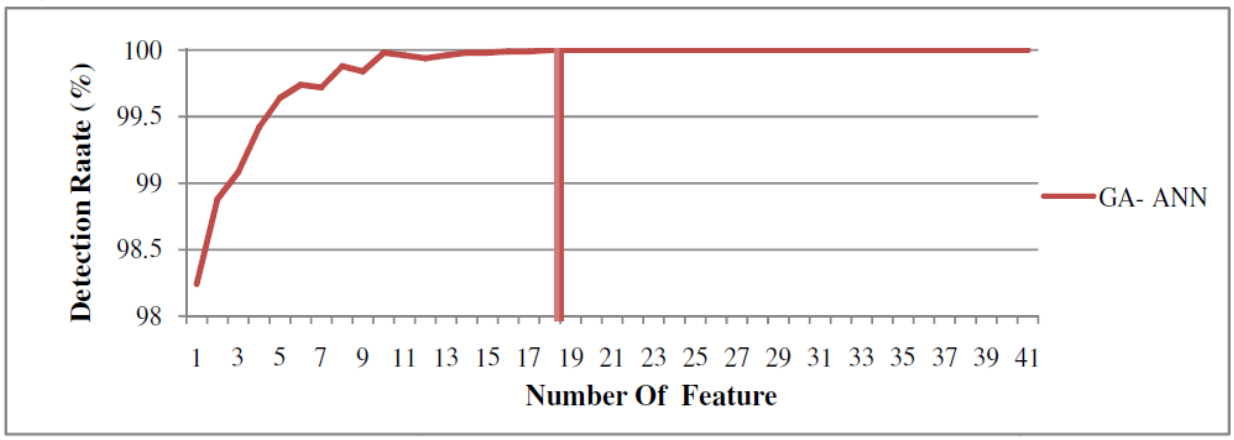

Fig 4: Result of detection rate for GA and ANN

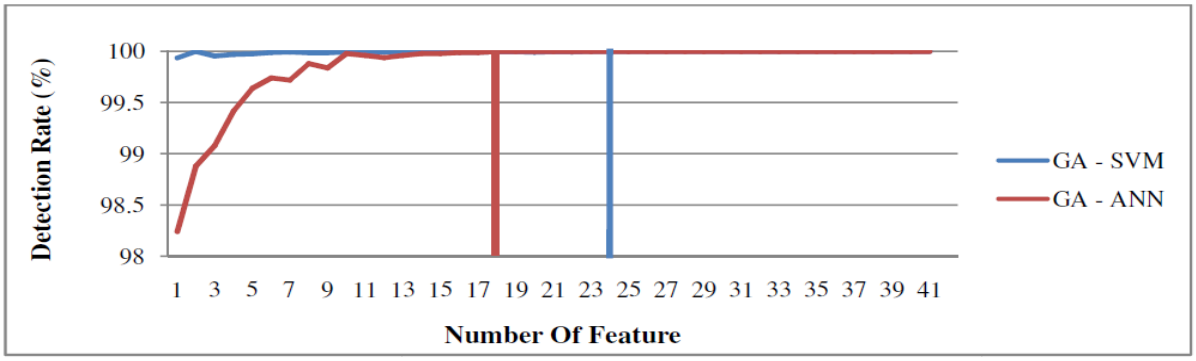

Fig 5: RESULT OF COMPARING ANN WITH GA AND SVM WITH GA

Table 2: COMPARATION OF GA ON ANN WITH OTHER ALGORITHM

Name of algorithm	Number of Feature
LCFS	**21**
FFSA	**21**
FUZZY RULE	**41**
SVM WITH GA	**24**
ANN WITH GA	**18**

8. CONCLUSION

It was noted that the methods available to detect intruders who still face the major issue of the high rate of false alarms. The idea at the basis of this work is to combine the advantages of both techniques to increase the precision of the intrusion detection system and also provide a false alarm rate. Therefore, it conducted a study to determine the best methods to undergo hybridization for intrusion detection system. It was noted that the GA is the best choice for the optimization problem. ANN also usually involves different detection rates so the idea at the basis of this investigation is to create a hybrid method that combines the selection function that is based on the GA with the classification of ANN to improve the system performance by providing high rate of detection & lower false negative.

9. REFERENCES

[1] J., Muna. M. and Mehrotra M., "Intrusion Detection System : A design perspective", Proceeding of 2rd International Conference On Data Management, IMT Ghaziabad, India.,2009,265-372.
[2] M. Panda, and M. Patra, “Building an efficient network intrusion detection model using Self Organizing Maps”, Proceeding of world academy of science, engineering and technology, 38, 2009, 22-29.
[3] M. Khattab Ali, W. Venus, and M. Suleiman Al Rababaa, "The Affect of Fuzzification on Neural Networks Intrusion Detection System", IEEE computer society,2009, 1236-1241.
[4] B. Mykerjee, L. Heberlein T., and K. Levitt N., "Network Intrusion Detection", IEEE Networks, 8(3), 1994, 14-26.
[5] W. Jung K., "Integration Artificial Immune Algorithms for Intrusion Detection", dissertation in University of London, 2002, 1-5.

[6] A. Chittur, "Model Generation for an Intrusion Detection System Using Genetic Algorithms", Technical Report, Ossining, New York, 2001.
[7] W. Li, "Using Genetic Algorithm for Network Intrusion Detection", http://www.security.cse.msstate.edu, Department of Computer Science and Engineering, Mississippi State University, USA, 2004.
[8] W. Lu, I. Traore, "Detecting New Forms of Network Intrusion Using Genetic Programming", Computational Intelligence, 20(3), Blackwell Publishing, Malden, 2004, 475-494.
[9] M. M. Pillai, J. H. P. Eloff, H. S. Venter, "An Approach to Implement a Network Intrusion Detection System using Genetic Algorithms", Proceedings of SAICSIT, 2004, 221-228.
[10] S. M. Bridges, R. B. Vaughn, "Fuzzy Data Mining And Genetic Algorithms Applied To Intrusion Detection", Proceedings of 12th Annual Canadian Information Technology Security Symposium, 2004, 109-122.
[11] J. Gomez, D. Dasgupta, "Evolving Fuzzy Classifiers for Intrusion Detection", Proceedings of the IEEE, 16(6), 2002, 1462-1475.
[12] R. H. Gong, M. Zulkernine, P. Abolmaesumi, "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection", IEEE, 2005, 246-253.
[13] B. Abdullah, I. Abd-alghafar, Gouda I. Salama, A. Abd-alhafez, "Performance Evaluation of a Genetic Algorithm Based Approach to Network Intrusion Detection System", Proceeding of 13th International Conference on AEROSPACE SCIENCES & AVIATION TECHNOLOGY, 2009, 1-17.
[14] M. Vallipuram and B. Robert, "An Intelligent Intrusion Detection System based on Neural Network", Proceeding of International Conference on Applied Computing, 2004, 356-362.
[15] R. H. Gong, M. Zulkernine, P. Abolmaesumi, "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection", Proceeding of 6th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2005, 246-253.
[16] M. Al-Subaie, "The power of sequential learning in anomaly intrusion detection", master degree thesis, Queen University, Canada, 2006.
[17] P. Kukielka and Z. Kotulski, "Analysis of different architectures of neural networks for application in intrusion detection systems", proceeding of the international conference on computer science and information technology, 2008, 807-811.
[18] M. Moradi and M. Zulkernine, "A Neural Network based system for intrusion detection and classification of attacks", submitted at Queen University, Canada, 2004, 148-154.
[19] D. Novikov, V. Roman Yampolskiy, and L. Reznik, "Artificial Intelligence Approaches For Intrusion Detection", Proceeding of Systems, Applications and Technology Conference, IEEE Long Island, 2006, 1-8.
[20] S. Lília de Sá, C. Adriana Ferrari dos Santos, S. Demisio da Silva, and A. Montes, "A Neural Network Application for Attack Detection in Computer Networks", Proceeding of IEEE joint conference on Neural Networks, 2, 2004, 1569-1574.
[21] P. Kukiełka and Z. Kotulski, "Analysis of Different Architectures of Neural Networks for Application in Intrusion Detection Systems", Proceedings of the International Multi conference on Computer Science and Information Technology, IEEE, 2008, 807– 811.
[22] KDD Cup 1999: Data; http://www.kdd.org/kddcup/index.php?section=1999&method=data