Effective Anomaly Intrusion Detection System based on Neural Network with Indicator Variable and Rough set Reduction

Rowayda A. Sadek[1,2] , M. Sami Soliman[3] and Hagar S. Elsayed[4]

[1] Information Technology Dept., Faculty of Computer Science and Information, Helwan University, Cairo, Egypt

[2] Computer Engineering Dept., Faculty of Eng., Arab Academy for Science and Technology & Maritime Transport, Cairo, Egypt

[3] National seismic network laboratory. Earthquake Dept., National Research Institute of Astronomy & Geophysics, Helwan, Egypt

[4] National seismic network laboratory. Earthquake Dept., National Research Institute of Astronomy & Geophysics, Helwan, Egypt

Abstract

Intrusion detection system (IDS) is an important tool for the defense of a network against attacks. It monitors the activities occurring in a computer system or network and analyzes them for recognizing intrusions to protect the computer network. Most of the existing IDSs use all of the 41 features available in the network packet to analyze and look for intrusive pattern, while some of these features are redundant and irrelevant. The weakness of this approach is the time-consuming during detection process and degrading the performance of IDSs. A well-defined feature selection algorithm makes the classification process more effective and efficient. In this paper a new hybrid algorithm NNIV-RS (Neural Network with Indicator Variable using Rough Set for attribute reduction) algorithm is used to reduce the amount of computer resources like memory and CPU time required to detect attack. Rough Set Theory is used to select out feature reducts. Indicator Variable is used to represent dataset in more efficient way. Neural network is used for network traffic packet classification. Tests and comparison were done on NSL-KDD dataset which is the improved version of KDD99 data set. The experiments results showed that the proposed algorithm gives better and robust representation of data as it was able to select features resulting in 80.4% data reduction, select significant attributes from the selected features and achieve detection accuracy about 96.7% with a false alarm rate of 3%.

Keywords: *intrusion detection, feature selection, indicator variable, neural network, NSL-KDD.*

1. Introduction

Network Security is very important as networks are exposed to an increasing number of security threats. An intrusion detection system is a software tool that monitors network or computer system for malicious activities [1]. Intrusions are defined as attempts to compromise the confidentiality, integrity or availability of a computer or network. They are caused by attackers accessing a system from the Internet, by authorized users of the systems who attempt to gain additional privileges for which they are not authorized or by authorized users who misuse the privileges given to them [2].

IDS aim to recognize unusual access or attacks to secure internal networks [3]. Network-based IDS is a valuable tool for the defense-in-depth of computer networks. It looks for known or potential malicious activities in network traffic and raises an alarm whenever a suspicious activity is detected [4].

Generally, there are two types of approaches for intrusion detection system: Misuse Detection and Anomaly Detection. Misuse detection stores the signatures of known attacks in the database and compares new instances with the stored signatures to detect attacks. The drawback of this approach is that it cannot detect new attacks and any new attack signatures have to be added manually in the list of known patterns. Anomaly Detection on the other hand studies the normal behavior of the monitored system and then looks out for any difference in it to detect intrusions so it is able to detect new attacks as any attack is assumed to be different from normal activity. However anomaly detection sometimes sets false alarms because it erroneously classifies the normal user behaviors as attacks [5].

The purpose of this research is to make anomaly intrusion detection system feasible with high detection accuracy and low false alarm rate.

In the experiment, NSL-KDD data set is used for the evaluation of the IDS which is an improved version of KDD99 data set [6]. It has solved some of the inherent problems of the KDD99 data set [7]. NSL-KDD data set

consists of selected records of the complete KDD99 data set. Each NSL-KDD connection record contains 41 features and is labeled as either normal or attack type.
In this work, we aim to filter out redundant, worthless information, which leads significantly to reduce the amount of computer resources, both memory and CPU time, required to detect attacks. We design a NNIV-RS (Neural Network with Indicator Variable using Rough Set for attribute reduction) algorithm based on feature selection and classification to enhance the detection accuracy of the IDS. Rough Set theory is used as a feature selection tool to select the most significant features, indicator variable is used to convert selected features into separated attributes and then RST is used again to select the important attributes for the IDS. feed forward propagation neural network is used as a classification tool to classify normal and different types of attacks.

The rest of the paper is organized as follows. Section 2 describes some related work for this research based on feature reduction and back propagation neural network. Section 3 describes the NSL-KDD dataset. Section 4 describes the detail analysis of the proposed algorithm. Section 5 describes experiments and results followed by a conclusion in Section 6.

2. Related Work

Some Researches were designing algorithms for intrusion detection system based on feature reduction tools and back propagation neural network classifier but they don't mention the effect of the reduction of features on the results of detection accuracy and false alarm rate.
Rupali Datti et. al. [8] proposed Linear Discriminant Analysis (LDA) to reduce features on the NSL-KDD dataset to 4 features only this gives 97% reduction in the input data and approximately 94% reduction in the training time. Shilpa Lakhina et. al. [9] proposed Principal Component Analysis (PCA) as a reduction tool, it reduces the features to 8 features this gives 80.4% data reduction and approximately 35%-40% reduction in the training time and 75%-80% reduction in the testing time. Neveen I. Ghali [10] proposed rough set theory (RST) to select features of the KDD99 dataset. Only 7 features are selected resulting in 83% reduction in the input data and 85%-90% time reduction and approximately 90% reduction in mean squared error in detecting new attacks.

3. NSL-KDD Dataset Description:

During the last decade, KDD-99 data set was the mostly widely used data set for the evaluation of the anomaly IDS. However, recent studies show that there are some inherent problems present in KDD-99 data set which highly affects the performance of evaluated systems, and results in a very poor evaluation of anomaly detection approaches. To solve these issues, a new data set, NSL-KDD is proposed, which consists of selected records of the complete KDD-99 data set [6].
NSL-KDD data set has the following advantages over the original KDD-99 data set:

- Train set is free from redundant records, so the classifiers will not be biased towards more frequent records.
- The number of selected records from each difficulty level group is inversely proportional to the percentage of records in the original KDD data set. As a result, the classification rates of distinct machine learning methods vary in a wider range, which makes it more efficient to have an accurate evaluation of different learning techniques.
- The numbers of records in the train and test sets are reasonable, which makes it affordable to run the experiments on the complete set without the need to randomly select a small portion. Consequently, evaluation results of different research works will be consistent and comparable.

Each NSL-KDD connection record contains 41 features as shown in Table 1 which have either continuous or discrete values (e.g., protocol type, service, flag, etc.) and it is labeled as either normal or an attack type. The simulated attacks fall in one of the Following categories [6]:

(i) Denial of Service Attack (DoS): is an attack in which the attacker makes some computing or memory resource too busy or too full to handle legitimate requests, or denies legitimate users access to a machine. For example: ping of death and SYN flood.
(ii) Probing Attack: the attacker scans a network of computers to gather information and then uses it to exploit the system. For example: Port scanning.
(iii) Remote to Local Attack (R2L): occurs when an attacker who does not have an account on a remote machine sends packet to that machine over a network and exploits some vulnerability to gain local access. For example: password guessing.
(iv) User to Root Attack (U2R): in this an attacker starts out with access to a normal user account on the system and is able to exploit vulnerability to gain root access to the system. For example: buffer overflow attack.

Table 1: 41 features of the NSL-KDD dataset

Feature No	Feature name	Type	Feature No	Feature name	Type
1	Duration	Con.	22	is_guest_login	Dis.
2	protocol_type	Dis.	23	Count	Con.
3	Service	Dis.	24	srv_count	Con.
4	Flag	Dis.	25	serror_rate	Con.
5	src_bytes	Con.	26	srv_serror_rate	Con.
6	dst_bytes	Con.	27	rerror_rate	Con.
7	Land	Dis.	28	srv_rerror_rate	Con.
8	wrong_fragment	Con.	29	same_srv_rate	Con.
9	Urgent	Con.	30	diff_srv_rate	Con.
10	Hot	Con.	31	srv_diff_host_rate	Con.
11	num_failed_logins	Con.	32	dst_host_count	Con.
12	logged_in	Dis.	33	dst_host_srv_count	Con.
13	num_compromised	Con.	34	dst_host_same_srv_rate	Con.
14	root_shell	Con.	35	dst_host_diff_srv_rate	Con.
15	su_attempted	Con.	36	dst_host_same_src_port_rate	Con.
16	num_root	Con.	37	dst_host_srv_diff_host_rate	Con.
17	num_file_creations	Con.	38	dst_host_serror_rate	Con.
18	num_shells	Con.	39	dst_host_srv_serror_rate	Con.
19	num_access_files	Con.	40	dst_host_rerror_rate	Con.
20	num_outbound_cmds	Con.	41	dst_host_srv_rerror_rate	Con.
21	is_host_login	Dis.	-	-	-

4. Proposed Algorithm

Proposed Intrusion detection algorithm is represented in figure1, which contains six steps. These steps are descried in this section. First, Data preprocessing is done to convert the non-numeric value to numeric value. After that, a Normalization process is performed on the numeric value to make it in the same range, and then feature selection method is used to select the most significant features using Johnson's Algorithm of RST. Then indicator variable is applied to the selected features to convert them into separated attributes. After that attribute selection is applied to the selected features to select the relevant attributes for data classification using Johnson's Algorithm of RST. Lastly, the reduct sets are sent to Feed forward Back-propagation neural network for classification of reduced dataset.

4.1 Data Preprocessing

Preprocessing is needed to convert the non-numeric value to numeric value because neural network classification uses only numerical data for training and testing.
Two steps are done in the preprocessing of the NSL-KDD data set:

1- Convert the non-numeric features to numeric value. In the NSL-KDD data set, all features of the data set take numeric values except three, namely, protocol type, service, and flag. Those three features will be converted to numeric value (e.g. for protocol type: TCP=3, UDP=7, ICMP=9).
2- Convert Attack names to its category: 0 for Normal, 1 for DoS (Denial of service), 2 for probe, 3 for R2L (remote to-local) and 4 for U2R (user-to-root).

4.2 Normalization

Since features of the NSL-KDD data set have either discrete or continuous values, the ranges of the features value were different and this made them incomparable. As a result, the features were normalized by using min-max normalization [11] to map all the different values for each feature to [0, 1] range.

4.3 Feature selection for IDS dataset

Subsequent to preprocessing of data, the features of the data set are identified as either being significant to the intrusion detection process, or redundant. This process is known as feature selection. Redundant features are generally found to be closely correlated with one or more other features. As a result, omitting them from the intrusion detection process does not degrade classification accuracy. In fact, the accuracy may improve due to the

resulting data reduction, and removal of noise and measurement errors associated with the omitted features [12].

Feature selection aims to reduce the number of irrelevant and redundant features of the intrusion data set to improve the classification detection accuracy. Moreover, Effective features selection is very important for constructing a high performance IDS.

Rough set theory (RST) is a useful mathematical tool to deal with imprecise and insufficient knowledge, find hidden patterns in data, and reduce dataset size [13]. Also, it is used for evaluation of significance of data and easy interpretation of results. RST contributes immensely to the concept of reducts. A reduct is a minimal subset of features with the same capability of objects classification as a whole set of features [10].

The following definitions as given in [13] show the reduct derivation for Rough set theory.

Definition 1:

Knowledge is represented by means of a table called an Information System given by S =<U,A,V,f >; where U = $\{x_1, x_2, \ldots, x_n\}$ is a finite set of objects of the universe (n is the number of objects) , A=$\{a_1, a_2, \ldots, a_m\}$; V= $\cup_{a\in A} V_a$ and V_a is a domain of feature a; $f: U\times A \rightarrow A$ is a total function such that $f(x,a)\in V_a$ for each $a\in A$, $x\in U$. If the features in A can be divided into condition set C and decision feature set D; i.e. $A=C\cup D$ and $C\cap D=\Phi$. The information system A is known as decision system or decision table.

Definition 2:

Every $B\subseteq A$ yields an equality relation up to indiscernibility, $IND_A(B)\subseteq (U\times U)$, given by: $IND_A(B) = \{(x,x') : \forall\, a\in B \;\; a(x) = (x')\}$ a reduct of A is the least $B\subseteq A$ that is equivalent to A up to indiscernibility. i.e., $IND_A(B) = IND_A(A)$.

To reduce the features of NSL-KDD data set we have used Johnson's Algorithm of RST for feature selection [14] which implements a variation of a simple greedy search algorithm. This algorithm extracts a single reduct only. After applied the Rough set theory algorithm, only 8 features is selected instead of large 41 features. This gives 80.4% reduction in input data.

4.4 Indicator variables method

Indicator variable is used to convert features into separated attributes.

The basic procedure of indicator variables (IV) is that, 1 indicates the occurrence of categories of features and 0 indicates its nonoccurrence of categories of features [15]. Indicator variable is applied to convert features into separated attributes.

Two steps are done during this stage

1- For discrete features IV is applied directly. (e.g. for protocol type: TCP, UDP, ICMP). If the connection type is tcp then indicator variable replace it with 1 and others with zero [1 0 0].

2- For continuous features it must be converted to discrete before applying IV therefore, k-mean clustering algorithm is used which group's objects based on their feature values into K disjoint clusters. Objects that are classified into the same cluster have similar feature values [16]. Each continuous feature (e.g. src_byte) takes 5 interval, so each value is replaced with a vector of 5 elements all with 0 value except the element that represents its cluster will be 1.

4.5 Attribute Selection

After successfully applying the indicator variable to the selected eight features, the dimensionality of attributes is increased to 95 attributes. To handle this problem Johnson's Algorithm of RST is used again to select the significant attributes for the Intrusion Detection System in order to improve the prediction ability of the classifier. 40 IV attributes are selected from the 95 attributes.

4.6 Classification using Feed Forward back propagation Neural Network

The goal of using artificial neural network (ANN) for intrusion detection is to be able to generalize from incomplete data and to be able to classify data as being normal or attack. An artificial neural network consists of a collection of processing elements that are highly interconnected and transform a set of inputs to a set of desired outputs. The result of the transformation is determined by the characteristics of the elements and the weights associated with the interconnections among them. By modifying the connections between the nodes the network is able to adapt to the desired outputs [17]. The ability of high tolerance for learning-by-example makes neural networks flexible and powerful in IDS [18].

Multi-Layer Feed Forward neural network consists of an input layer, one or more hidden layers and an output layer of neurons. Every node in a layer is connected to every other node in the neighboring layer. All of the connections are in a forward direction only. That is why such a structure is known as fully connected, feed-forward, multilayer network [19]. An example of feed forward neural network with one hidden layer is shown in Figure2.

The most popular and widely used learning algorithm for multilayer feed forward Neural Networks is the back propagation algorithm. It is based on the Delta Rule that basically states that if the difference (delta error) between the user's desired output (target) and the network's actual output is to be minimized, the weights must be continually modified. The result of the transfer

function changes the delta error in the output layer. The error in the output layer has been adjusted, and therefore it can be used to change the input connection weights so that the desired output may be achieved. The learning mechanism is illustrated in Figure3 [20].

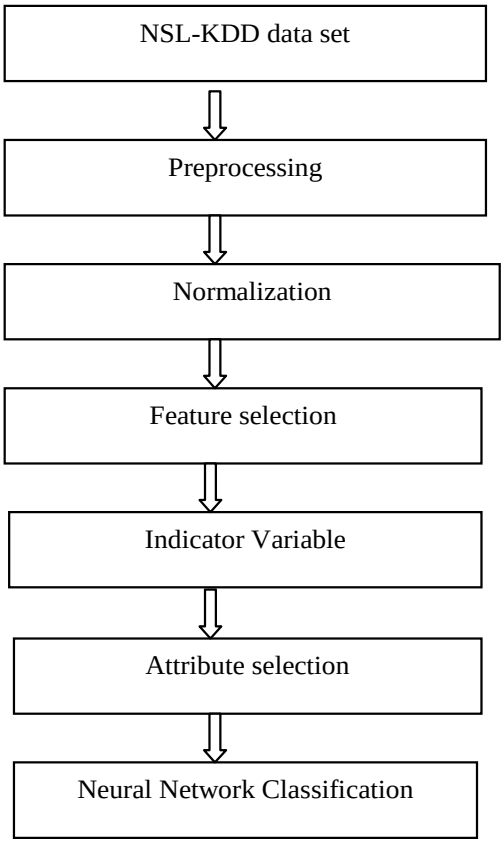

Figure1: flowchart of the proposed IDS

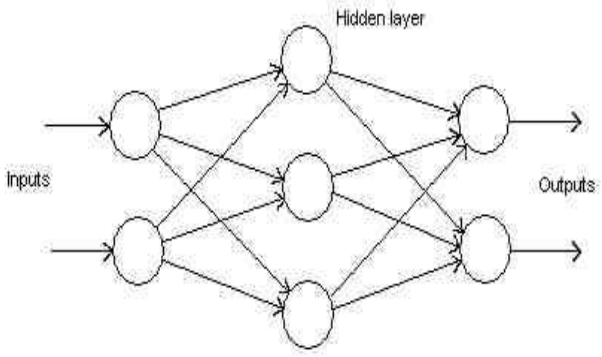

Figure2: Feed forward neural network

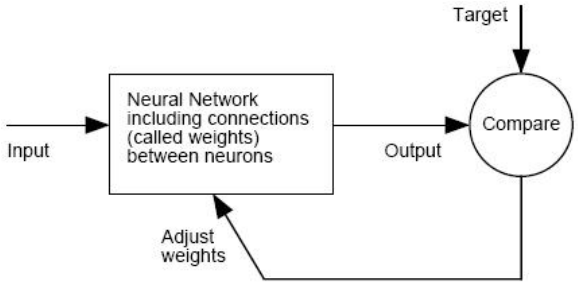

Figure.3: Neural network learning mechanism

5. Experiments and Results

The experiments were run on a system with a 3.10GHZ core i5 processor and 6GB of RAM running windows 7. All the processing is done using MATLAB® 2010b. MATLAB's Neural Network Toolbox was used for designing a feed forward back propagation neural network, whereas rough set operations were done in ROSETTA. It is a data mining tool developed by Ohrn [21]. The algorithm used by ROSETTA library supports two categories of discrenibility:
1). full: reducts are selected relative to the system as a whole in this category of discrenibility.
2). Objects: reducts are selected relative to a single object In this category of discrenibility.
By applying the feature selection Johnson algorithm, we get the reduct set. The resulting selected features are only 8, This gives 80.4% reduction in input data then Johnson's Algorithm is used again after applying the indicator variable to select the significant attributes for classification.

Following fundamental formulas are used to evaluate the performance of the system: The detection accuracy rate and the false alarm rate were calculated according to the following assumptions [22]:

- False Positive (FP): the total number of normal records that are classified as anomalous
- False Negative (FN): the total number of anomalous records that are classified as normal
- Total Normal (TN): the total number of normal records
- Total Attack (TA): the total number of attack records
- Detection Rate = [(TA-FN) / TA]*100
- False Alarm Rate = [FP/TN]*100

NSL-KDD training dataset is used to train and test intrusion detection system (16799 record of training sample and 9000 record of test samples were used). A MATLAB feed forward neural network program with 10 hidden layer and 5 output neurons is used to evaluate the performance of the intrusion detection system in terms of detection accuracy and false alarm rate.

The results along with the comparison to other existing methods using NSL-KDD data set are shown in Table 2.

Table 2: Detection Accuracy Comparison of Machine learning algorithms using NSL-KDD Dataset

Classifier	**Detection Accuracy (%)**	**False Alarm Rate in %**
AdaBoost [23]	90.31	3.38
Discriminative Multinomial Naïve Bayes +PCA [24]	94.84	4.4
Discriminative Multinomial Naïve Bayes +RP [24]	81.47	12.85
Discriminative Multinomial Naïve Bayes +N2B [24]	96.5	3.0
Enhanced Resilient Backpropagation Artificial Neural Network [25]	94.7	----
Learning Vector Quantization and an enhanced resilient backpropagation artificial neural network [26]	97.06	----
Neural Network with Indicator Variable using Rough Set for attribute reduction (proposed algorithm)	**96.7**	**3.0**

6. Conclusion

This research focuses on the effect of dimensionality reduction using attributes selection on building effective intrusion detection system with high detection accuracy and low false alarm rate. The proposed algorithm NNIV-RS (Neural Network with Indicator Variable using Rough Set for attribute reduction) algorithm is used for this purpose. Experimental results showed that the proposed algorithm gives better and robust representation of data as it was able to reduce the number of features resulting in 80.4% reduction in input data and it was able select significant attributes which leads to improve the detection accuracy to 96.7% with a false alarm rate of 3%.The results showed that the proposed algorithm is reliable and efficient in intrusion detection.

7. References

[1] K. Scarfone, P. Mell, “Guide to Intrusion Detection and Prevention Systems (IDPS)”. Computer Security Resource Center (National Institute of Standards and Technology). February 2007.

[2] Akbar, Shaik, K. Nageswara Rao, and J. A. Chandulal. "Intrusion detection system methodologies based on data analysis." *International Journal of Computer Applications IJCA* 5.2, 2010: 10-20.

[3] C. Tsai , Y. Hsu, C. Lin and W. Lin, ”Intrusion detection by machine learning: A review”, Expert Systems with Applications, vol. 36, pp.11994-12000, 2009.

[4] Govindarajan, M., and R. M. Chandrasekaran. "Intrusion Detection using an Ensemble of Classification Methods." Proceedings of the World Congress on Engineering and Computer Science. Vol. 1. 2012.

[5] Rawat, S., Gulati V., Pujari A., A Fast Host-based Intrusion Detection System Using Rough Set Theory, Transaction on Rough Sets IV, LNCS 3700, 2005.

[6] Tavallaee, Mahbod, et al. "A detailed analysis of the KDD CUP 99 data set." Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defence Applications, 2009.

[7] KDD Cup 1999. Available on http://kdd.ics.uci.edu/ Databases/kddcup 99/kddcup99.html,Ocotber 2007.

[8] Rupali Datti,Bhupendra Verma,"Feature Reduction for Intrusion Detection Using Linear Discriminant Analysis." (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 04, 2010, 1072-1078.

[9] Shilpa Lakhina, Sini Joseph and Bhupendra Verma, “Feature reduction using using Principal Component Analysis for Effective AnomalyBased Intrusion Detection on NSL-KDD”, Int. J. of engineering science and technology, Vol. 2(6), 2010, 1790-1799.

[10] Neveen I. Ghali, "Feature selection for effective anomaly based intrusion detection." IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.3, March 2009.

[11] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques: Concepts and Techniques. Elsevier, 2011.

[12] Zubair A. Baig, Abdulrhman S. Shaheen, and Radwan AbdelAal, "One-Dependence Estimators for Accurate Detection of Anomalous Network Traffic," International Journal for Information Security

Research (IJISR), Volum 1, Issue 4, December 2011.

[13] Pawlak, Z., Rough Sets: "Theoretical Aspects of Reasoning about Data", Kluwer Academic Publishers, 1991.

[14] Godinez, F., Hutter, D., Monroy R., Attribute Reduction for Effective Intrusion Detection, AWIC 2004, LNAI 3034, 2004.

[15] Munawar, Saima, Mariam Nosheen, and Haroon Atique Babri. "Anomaly Detection through NN Hybrid Learning with Data Transformation Analysis",International Journal of Scientific & Engineering Research Volume 3, Issue 1, January 2012.

[16] Gerhard Münz, Sa Li, and Georg Carle, " Traffic anomaly detection using k-means clustering" , In Proceedings of Leistungs-,Zuverlässigkeits-und Verlässlichkeitsbewertung von Kommunikationsnetzen und Verteilten Systemen, GI/ITG-Workshop MMBnet, September 2007.

[17] Sun, Ning-Qing, and Yang Li. "Intrusion detection based on back-propagation neural network and feature selection mechanism." Future Generation Information Technology. Springer Berlin Heidelberg, 2009. 151-159.

[18] Liang-Bin, Lai, Chang Ray-I, and Kouh Jen-Shiang. "Detecting network intrusions using signal processing with query-based sampling filter." EURASIP Journal on Advances in Signal Processing, 2009.

[19] Akerkar, Rajendra, and Priti Sajja. Knowledge-based systems. Jones & Bartlett Publishers, 2010.

[20] Alsaleh, Omar I., and Abdulkader A. Alfantookh. "Improved detection system of denial of service attack." 3rd IEEE Gulf International Conference (IEEE-GCC 2007). 2007.

[21] Ohrn, A., Komorowski, J., A Rough Set Toolkit for Analysis of Data, In Proceedings of the third Joint conference on Information Sciences, Vol(3), USA, 1997, pp.403-407,available on http://www.idi.ntnu.no/~aleks/rosetta.

[22] S.T. Sarasamma, Q.A. Zhu, and J. Huff, "Hierarchal Kohonenen Net for Anomaly Detection in Network Security," IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 35(2), 2005, pp. 302-312.

[23] V.P. Kshirsagar and Dharmaraj R. Patil, "Application of Variant of AdaBoost based Machine Learning Algorithm in Network Intrusion Detection", International Journal of Computer Science and Security (IJCSS), Vol. 4, Issue.2, 2010, pp. 1-6.

[24] Panda, M., Abraham, A., Patra, M.R., "Discriminative multinomial Naïve Bayes for network intrusion detection", Information Assurance and Security (IAS), 2010 Sixth International Conference, 2010, pp. 5-10.

[25] Naoum, Reyadh Shaker, Namh Abdula Abid, and Zainab Namh Al-Sultani. "An Enhanced Resilient Backpropagation Artificial Neural Network for Intrusion Detection System.". IJCSNS International Journal of Computer Science and Network Security, VOL.12 No.3, March 2012.

[26] Naoum, Reyadh Shaker, and Zainab Namh Al-Sultani. "Hybrid System Of Learning Vector Quantization And Enhanced Resilient Backpropagation Artificial Neural Network For Intrusion Classification." IJRRAS International Journal of Research and Reviews in Applied Sciences, vol14 issue2,2012.

AUTHORS PROFILE

Rowayda A. Sadek Associate Prof. Rowayda A. Sadek received her PhD. in 2005 from Alexandria University, Alexandria, Egypt in Communication and Electronics Engineering. She is currently working as Associate Prof. in Information Technology Department, Faculty of Computers and Information, Helwan University, Cairo, Egypt. She worked as Assoc. Prof. in Computer Engineering Department, in the College of Engineering & Technology in AASTMT. Her research interests include Computer Networking and Security, Multimedia Processing for image, audio, video, etc. also Multimedia Networking, and Security as interdisciplinary research.

M. Sami Soliman has received his B.Sc of Computer Engineering and Automatic control from Tanta University, Egypt in 2001. Master of Computer Engineering and science from Menofia University , Egypt in 2007 and PhD of Computer Engineering from Central South university ,China in 2010. he is currently an expert in Saudi National Center for earthquakes and volcano.

Hagar S. Elsayed has received her B.Sc of computer engineering from october 6 university, Egypt. Currently working for master degree in Arab Academy for Science and Technology & Maritime Transport.