Intrusion Detection Using PCA Based Modular Neural Network

Khaled Al-Nafjan , Musaed A. Al-Hussein , Abdullah S. Alghamdi, Mohammad Amanul Haque, and Iftikhar Ahmad

***Abstract*—Most of current intrusion detection systems are based on machine learning methods but very few till now use clustering algorithms as a preprocessing layer to reduce the high dimensionality of data, which is difficult to analyze. In this paper we introduce Modular Neural Network for intrusion detection, which apply Principal Component Analysis (PCA) as preprocessing layer for reducing huge information quantity presented in knowledge discovery and data mining (KDD99) data set. PCA significantly reduce the high dimensionality of data set without loss of information. Then this preprocess data in the form of principal component is presented to Batch Backpropagation Neural Network for efficient intrusion detection. We rely on some experiments to calculate Root Mean Square Error (RMSE) using Modular Neural Network on KDD 99 data set. Our experimental results show improvement in the learning time due to the reduction of high dimensions of data. Also we have obtained low RMSE during training, which is below the acceptance range of 0.1. Proposed Modular Neural Network has capability to efficiently and accurately classify data into attack and normal.**

***Index Terms*—Intrusion Detection, Principal Component Analysis, Modular Neural Network, KDD99 dataset, Batch Backpropagation Neural Network.**

I. Introduction

Network Security is one of the key challenge faces by computer researcher presently. With the extensive use of computer networks, network security has become a most important concern for the developers and users of these networks. Consequently, the dilemma of intrusion detection has grasped the attention of research with the aim of deploying efficient intrusion detection systems (IDS). Currently available Intrusion detection mechanisms suffer with problem of high dimensional data; it is difficult to understand the underlying structure. Furthermore, the storage, transmission and processing of high dimensional data places great demands on systems. All these are aspects of one of the most interesting computational and data analysis problems [1].

Research in intrusion started about thirty years ago. James P. Anderson was considered to be the pioneer of Intrusion Detection [2]. James P. Anderson proposed certain types of threats to the security of computer systems could be identified through a review of information contained in the system's audit trail. Many operating systems automatically create a report which details the activity occurring on the system. In 1987 Dorothy Denning and Peter Neumann proposed a model, which is considered as a milestone in the field of Intrusion Detection. They define the foundation elements of Intrusion Detection [3]. IDS can be categorized in to Misuse detection and Anomaly detection. Further, IDS can also be classified into Host based, Network-based on the base of data source. Further, there are two techniques available, Statistical-Based Intrusion Detection (SBID) and Rule-Based Intrusion Detection (RBID) [4], [5].

The rest of paper is organized as follow: Section II discusses related work in Intrusion Detection using Artificial Neural Networks. Section III discusses artificial neural networks and their advantages in the area of intrusion detection. Section IV provides introduction to Principal Component Analysis and its applicability in Intrusion Detection. Sections V discuss in detail the Modular Neural Network. Section VI discusses the experiments and results. Finally Section VII provides a conclusion and recommendations for future work.

Manuscript received June 26, 2012; revised September 8, 2012.

The authors are with Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Kingdom of Saudi Arabia (e-mail: ghamdi@ksu.edu.sa; amanulhaque80@gmail.com; kalnafjan@ksu.edu.sa).

II. Related Work

IDS research has been ongoing actively for past three decade producing number of viable system. Some of which have been implemented on commercial level. But research in Intrusion Detection using Artificial Neural Networks started in last few years with great prospect. James Cannady was considered to be first to use artificial Neural Network for Intrusion detection. Before him rule base expert systems are frequently use. In his paper he explain the characteristic of Artificial Neural Network that Artificial neural networks provide the potential to identify and classify network activity based on limited, incomplete, and nonlinear data sources. He presents an approach to the process of misuse detection that utilizes the analytical strengths of neural networks, and provides the results from preliminary analysis of this approach. [6].

Jake Ryan et al discuss NNID; a backpropagation neural network called NNID (Neural network intrusion detector) was trained in the identification task and tested experimentally on a system of 10 users. The system was 96 % accurate in detecting unusual activity with 7 % false alarm rate. This suggests that learning user profile is an effective way for detecting intrusions. The NNID system works in three steps such as collecting data, training data and performance. If NN suggestion is different from the actual

user then indicate anomaly. If NN activation is greater than 0.5 then identification was correct otherwise less than 0.5 then anomalies are detected. It provides high degree of accuracy out of 24 intruders the network identified 22. It operates offline on daily logs not for real [7].

Rhodes et al uses Multiple SOMS the unsupervised learning to identify anomalies. He measures a 10 % difference between the measures of fit for the same vector on different runs. In case of data and particular distance measures all of the normal traffic scored between 0 and 3. Roughly 7 packets transmitted to accomplish the exploit, 2 registered just above 80, indicating they did not fit well on the map at all and 2 other registered above 630 indicating external anomaly. It can also be applied for analysis of data collected from network monitoring. The ratio of normal to intrusive packets was computed. The overflow is also detected by this NN. By learning to characterize normal behavior, it completely prepares itself to detect any abnormal network activity [8].

J Cannady uses CMAC (Cerebellar Model Articulation Controller) adaptive NN for intrusion detection that is capable of learning new attacks rapidly through the use of a modified reinforced learning method that uses feedback from the protected system. It provides online learning of attack patterns. It has rapid learning of data. It is extremely accurate in identify priori attack patterns. This modified reinforce learning approach resulted in an average error of 3.28^{-05} %, compared with an average error of 15 % in existing intrusion detection. The average error rate is 2.199 % that identify new attacks based on its experience [6].

Hammerstrom et al used ANNMD (Artificial Neural Network for Misuse Detection) a MLP (Multi Level Perceptron) architecture that consists of four connected layers with 9 inputs and 2 output nodes. The training of the neural network was conducted using a backpropagation algorithm for 10,000 iterations of the selected training data. Like the feed-forward architecture of the neural network, the use of a backpropagation algorithm for training was based on the proven record of this approach in the development of neural networks for a variety of applications. Of the 9,462 records, which were preprocessed for use in the prototype, 1000 were randomly selected for testing and the remaining were used to train the system. The training/testing iterations of the neural network required 26.13 hours to complete [9].

Morteza Amini et al. used UNNID (Unsupervised Neural Net based Intrusion Detector) system that detects network-based intrusions and attacks using unsupervised neural networks. Two types of unsupervised Adaptive Resonance Theory (ART) nets (ART-1 and ART-2) are used in this system. This system can efficiently classify network traffic into normal and intrusive. It has hybrid approaches of misuse and anomaly detection, so capable of detecting known attack types as well as new attack types as anomalies. The result of this system shows that ART-1 93.5 % of times and ART-2 90.7 % were able to recognize attack from normal one. This system has many advantages over previous nets because of its unsupervised learning. This capability improves its analysis of new data over time without the requirement of retraining over all the previous and new data [10].

III. Artificial Neural Networks

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 [11], [1], [4], [6].

A. *Artificial Neural Network in Intrusion Detection Systems*

A limited amount of research has been conducted on the application of neural networks to detecting computer intrusions. Artificial neural networks offer the potential to resolve a number of the problems encountered by the other current approaches to intrusion detection. Artificial neural networks are alternatives. Neural networks were specifically proposed to identify the typical characteristics of system users and identify statistically significant variations from the user's established behavior [1].

B. *Advantages of Neural Network-based Intrusion Detection Systems*

The first advantage in the utilization of a neural network in the detection would be the flexibility that the network would provide. A neural network would be capable of analyzing the data from the network, even if the data is incomplete or distorted. Similarly, the network would possess the ability to conduct an analysis with data in a non-linear fashion. Further, because some attacks may be conducted against the network in a coordinated attack by multiple attackers, the ability to process data from a number of sources in a non-linear fashion is especially important. [6], [4]. The inherent speed of neural networks is another benefit of this approach. Because the output of a neural network is expressed in the form of a probability the neural network provides a predictive capability to the detection of instances of misuse. A neural network-based misuse detection system would identify the probability that a particular event, or series of events, was indicative of an attack against the system. As the neural network gains experience it will improve its ability to determine where these events are likely to occur in the attack process. This information could then be used to generate a series of events that should occur if this is in fact an intrusion attempt. By tracking the subsequent occurrence of these events the system would be capable of improving the analysis of the events and possibly conducting defensive measures before the attack is successful [11].

However, the most important advantage of neural networks in misuse detection is the ability of the neural network to "learn" the characteristics of misuse and identifies instances. The probability of an attack against the system may be estimated and a potential threat flagged whenever the probability exceeds a specified threshold [1], [4].

IV. Principal Component Analysis

Principal component analysis (PCA) is a mathematical

procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The objective of principal component analysis is to reduce the dimensionality (number of variables) of the dataset but retain most of the original variability in the data [12]. Principal component analysis (PCA) has been called one of the most valuable results from applied linear algebra. PCA is used abundantly in all forms of analysis -from neuroscience to computer graphics - because it is a simple, non-parametric method of extracting relevant information from confusing data sets. With minimal additional effort PCA provides a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified structure that often underlie it [13].

PCA is wonderful multivariate statistical algorithm to reduce the different representation spaces before applying some machine learning algorithms on the different KDD99 Intrusion Detection datasets. This new representation permits to improve the learning time and space representation of the different datasets with a similar successful prediction in the whole experiments [12]. Below Fig .1. shown block diagram of modular neural network.

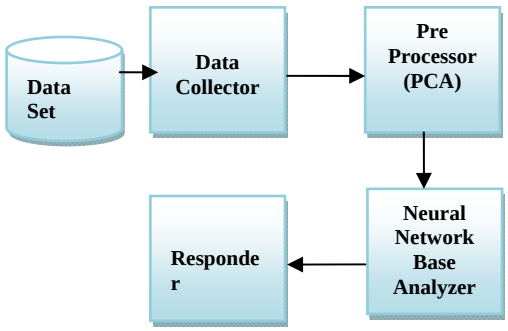

Fig. 1. Block diagram of modular neural network

V. Modular Neural Network

Our proposed Modular Neural Network has two layers. First layer is preprocessing layer, which use Principal Component Analysis (PCA) for clustering the input data. The main aim of using PCA is to reduce high dimension of KDD 99 data set [14], [15].

Second layer consist of main Neural Network, which uses Batch Backpropagation algorithm for intrusion detection. The input to main neural network is preprocessed input given by PCA.

A. *Data Collector*

This component collects data from audit data record. In case of our experimental result we are using KDD 99 data set, which is considered as trademark for research in intrusion detection.

B. *Preprocessor*

The Preprocessor component gets data from Data Provider and clusters it using PCA to reduce high dimensions. Steps of PCA are explained by diagram below.

C. *Neural Net based Analyzer*

The main component of Modular Neural Network is Neural Net based Analyzer, which analyzes the input given by Preprocessor and detects intrusions and attacks. It uses Batch Back Prorogation Algorithm A supervised Multi layers Feed forward Neural Network [6]. First we have to train it by some portion of KDD 99 data set until we got RMSE in acceptable range. Then test it for appropriate results in output file. Below Fig. 2. shown Steps involve in principal component analysis.

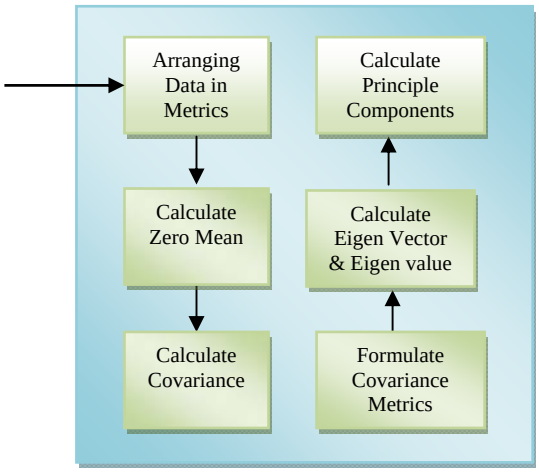

Fig. 2. Steps involve in principal component analysis

D. *Basic Steps of Backpropagation Algorithm are*

1) *Training Phase*

- Feed forward of the input training pattern
- Calculation and Backpropagation of associated error after specified batch size.
- Adjustment of the weights

2) *Testing Phase*

- Computation of the feed forward phase Responder

The detected intrusions by the Neural Network base analyzer are given to responder that classifies it as normal or attack.

VI. Experiments and Results

For experimental results we use JOONE (Java Object Oriented Neural Engine), for Implementing Modular Neural Network [16]. We use JoonePAD for Graphical representation of PCA principal components (clusters) and root mean square error (RMSE) during learning phase of batch back propagation. Fig. 4. is show Graphical representation of PCA output in Joone PAD. Fig. 5. is show property window of JOONE calculating RMSE during Training phase and Fig. 6.shown Graph of RMSE during training phase of modular neural network .

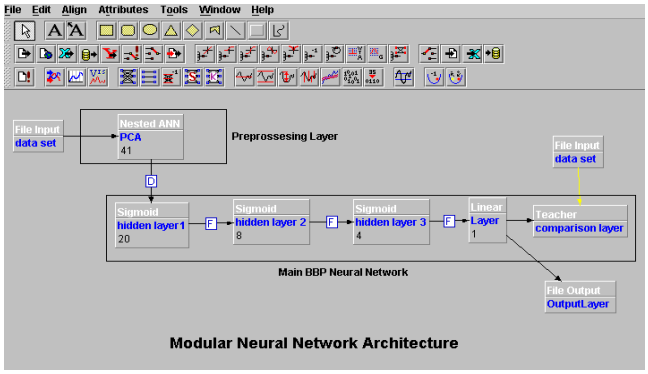

Fig. 3. Architecture of modular neural network in JOONE editor

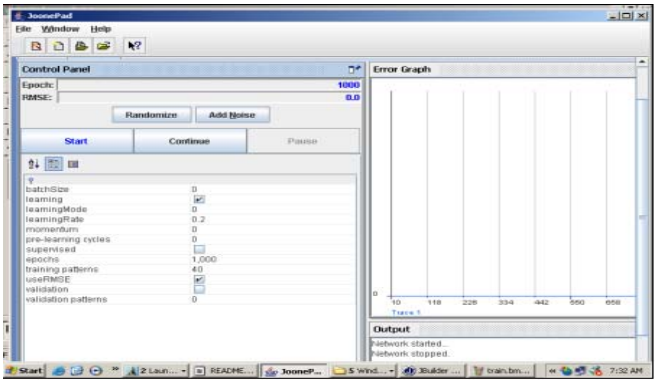

Fig. 4. Graphical representation of PCA output in Joone PAD

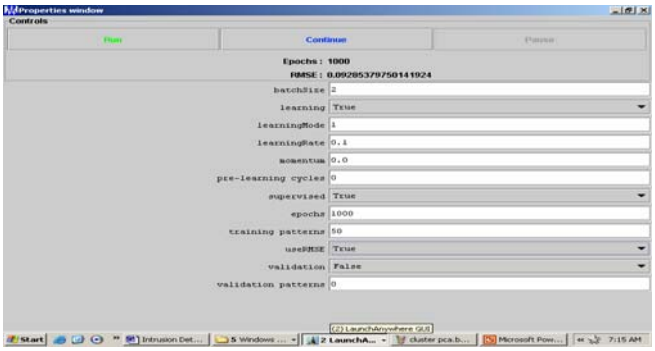

Fig. 5. Property window of JOONE calculating RMSE during training phase

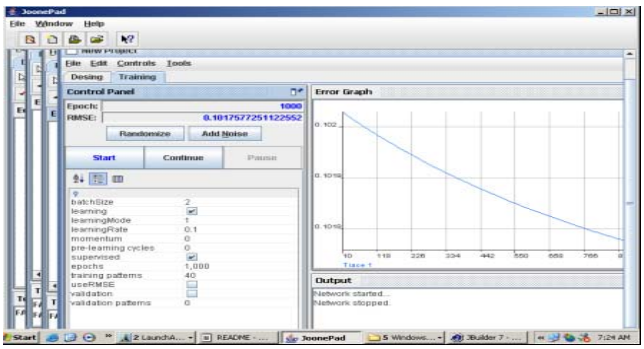

Fig. 6. Graph of RMSE during training phase of modular neural network

A. *Steps Involve in Implementing Modular Neural Network Using JOONE*

a) Build a PCA NN (by using the Sanger Synapse) and train it in unsupervised mode

b) Export it to a file in a serialized format (after having removed the i/o components used during the training).

c) Build the main neural network, and insert a Nested Neural Layer as first layers

d) Import the above-serialized PCA NN into the Nested Neural Layer

e) Set to false the 'learning' property of the Nested Neural Layer (it should already be set to that value by default)

f) Set to true the learning mode of the main NN

g) Randomize the weights of the main neural network

h) Start the training phase

i) Repeat the steps 7 and 8 until you get a good RMSE.

The RMSE we have got during training phase of Modular Neural Network is 0.09285379750141924, which is below 0.1, the acceptable range. Also the test data has detection rate of 96% for known attacks with very few false alarms. Though we are compromising on the integrity of input data set by reducing its high dimensions to low but still we got very good results.

VII. Conclusion and Future Work

We have presented in this research paper a new idea of how to reduce the size of data set using PCA before applying Neural Network learning algorithm Batch Backpropagation (BBP) on the KDD 99 intrusion detection datasets. This new approach permits us to improve the learning time and reduce the RMSE, which is below the acceptance value of 0.1. In future work we try to implement this Modular Neural Network in real time environment. Also try to detect the specific type of attacks.

References

[1] J. Cannady, "Artificial neural networks for anamoly detection," in *Proc. of National Information Systems Security Conference*, 2000, pp. 281–288.

[2] I. Ahmad, A. B. Abdullah, and A. S. Alghamdi, "Application of Artificial Neural Network in Detection of Probing Attacks," IEEE Symposium on Industrial Electronics and Applications (ISIEA 2009), October 4-6, 2009, Kuala Lumpur, Malaysia.

[3] J. P.Anderson "Computer Security Threat Monitoring and Surveillance, Technical Report, Company, Fort Washington, USA, 1980.

[4] J. Ryan, M. J. Lin, and R .Miikulainen. "Intrusion Detection with Neural Networks," *Advances in Neural Information Processing Systems* (1998). 10:943-949.

[5] B. C. Rhodes, A.M.J, and D.C.J., "Multiple Self-Organizing Maps for Intrusion Detection". *NIST National Information Systems Security Conference* (2000).

[6] D. Hammerstrom,"Neural Networks At Work.," IEEE Spectrum (1993). pp. 26-53.

[7] Y. Bouzida, F. Cuppens, and N. Cuppens-Boulahia , "Efficient Intrusion Detection Using Principal Component Analysis Gombault," *3rd Conférence sur la Sécurité et Architectures Réseaux* (SAR), La Londe, France. June 2004.

[8] A. R. Calvo, M. Partridge, and M. A. Jabri, "A Comparative Study of Principal Component Analysis Techniques," University of Sydney NSW, 2006.

[9] D. Dorothy , "An Intrusion-Detection Model," *IEEE Transactions on Software Engineering*, vol. SE-13, no.2.1987.

[10] M. Amini, R. Jalili, and H. R. Shahriari, "RT-UNNID: A practical solution to real-time network-based intrusion detection using unsupervised neural networks," Computers & Security (2006): 25(6), pp.459-468.

[11] I. Ahmad, A. B. Abdullah, and A. S. Alghamdi, "Evaluating Intrusion Detection Approaches Using Multi-criteria Decision Making Technique," Information Sciences and Computer Engineering(IJISCE), Australia, vol.1, no.1, pp. 60-67, 2010.

[12] I. Ahmad, A. B.Abdullah, and A. S. Alghamdi, "Application of Artificial Neural Network in Detection of DOS Attacks," in *Proceedings of the 2nd international Conference on ecurity of information and Networks* (2009). SIN '09. ACM, New York, NY, pp 229-234.

[13] Z. Sun, G. Bebis, and R. Miller, "Object detection using feature subset selection," *Pattern Recognition*, vol. 37, Issue 11, November 2004, pp. 2165-2176, DOI: 10.1016/j.patcog.2004.03.013.

[14] The 3rd International Knowledge Discovery and Data Mining Tools Competition, website link accessed on Janurary 2010, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[15] JOONE, [Online]. Available: http://www.joone.org

[16] I. Ahmad, A. B. Abdullah, and A. S. Alghamdi, "Artificial Neural Network Approaches to Intrusion Detection: A Review," *Telecommunications and Informatics conference*, Istanbul, Turkey, 2009, pp. 200-205.

Khalid Alnafjan is an assistant professor in software engineering department, college of computer and information sciences, King Saud University in Riyadh, Saudi Arabia. He was born in 1966 in Riyadh. Dr Khalid has obtained his bachelor degree from the same college above. He has obtained a master degree in software engineering and a PhD in computer science from the department of computer science, University of Sheffield, United Kingdom.

His major research interests lie in different areas including software quality assurance, C4I, and software engineering education. He has published several research articles in different conferences and academic journals Dr. Khalid is a member of Saudi computer society and British computer society.

Musaed A. Al-Hussein was born in Riyadh, Saudi Arabia. He received his B.S. degree in Computer Engineering from King Saud University, Riyadh, in 1988, and the M.S. and Ph.D. degrees in Computer Science and Engineering from University of South Florida, Tampa, Florida, in 1992 and 1997, respectively. Since 1997, he has been on the Faculty of the Computer Engineering Department, College of Computer and Information Science, King Saud University.

His current research interests include wireless sensor networks and wireless ad hoc networking.

Abdullah Alghamdi is a full time professor, SWE Department, College of Computer and Information Sciences, KSU. He holds a Ph.D. in the field of Software Engineering from the department of computer science, Sheffield University, UK, 1997. He obtained his M.Sc. in the field of software development technologies from the UK in 1993.

In the academic year 2004/5 he worked as a visiting professor at School of IT and Engineering, University of Ottawa, Ottawa, Canada, where he conducted intensified research in Web Engineering as part of his Post-Doc program. He recently published a number of papers in the field of Web engineering methodologies and tools.

Prof. Abdullah worked as a part-time consultant with a number of governmental and private organizations in the field of IT strategic planning and headed a number of IT committees inside and outside KSU. Currently he is chairing the Software Engineering Department and C4I CAS at KSU and part time consultant at Ministry of Defense and Aviation.

Mohammad Amanul Haque is lecturer in software engineering department, college of computer and information sciences, King Saud University, Riyadh, Saudi Arabia. He was born in India. Mohammad Amanul Haque has obtained his Master degree in Computer Science from Jamia Hamdard , New Delhi ,India in 2006 and Bachelor degree in Information Technology from Manipal University, India in 2003.

His research interests include Computer Networks, Network Security, Data-Mining, Quality Assurance and C4I systems. He has Microsoft Certification on DOT NET framework and member of IAENG.

Iftikhar Ahmad received the B.Sc. degree in Mathematics and Physics from Islamia University, Bahawalpur, Pakistan, in 1999 and the M.Sc. Computer Science from University of Agriculture, Faisalabad, Pakistan in 2001. He obtained his MS/M.Phil degree in Computer Science from COMSATS Institute Information Technology, Abbottabad, Pakistan in 2008. He received his Ph.D. degree in IT from Universiti Teknologi PETRONAS, Malaysia in 2011.

He has extensive experience of teaching different IT related subjects and network management. Moreover, he has different industrial certifications such as MCSE, CCNA, Linux (OS) and CCAI. He has published several papers in highly reputed international conferences and journals. He is member of IEEE, IAENG and IACSIT. His research interests include Computer Networks, Network Security, Intrusion Detection, Neural Networks, Analytic Hierarchy Process, and C4I systems.