Applying Hopfield Artificial Network and Simulating Annealing for Cloud Intrusion Detection

Bashair Al-Shdaifat, Wafa' Slaibi Alsharafat, Mohmmad el-bashir
AL al-Bayt University
Jordan
wafa@aabu.edu.jo

ABSTRACT: *Recently, Cloud Computing is a new paradigm trend in network environments that handles and manages a vast number of users by sharing services and data, for achieving this mission; safety and security of shared services and data are the main concerns for this type of environment. Intrusion Detection is an effective technique used to deal with security violations in such environment. Here, an Anomaly Intrusion Detection model of cloud environment will be proposed which based on using hybrid artificial intelligent algorithms; Hopfield neural networks simulated annealing.*

Keywords: Anomaly Detection, Cloud Environment

Received: February 25, 2015, Revised March 26 March 2015, Accepted 31 March 2015

1. Introduction

Cloud computing is classified as a distributed environment, which consists of a set of resources and services, in order to enable resource sharing and services in terms of scalability, and managing computing services, that are delivered upon request through the network [1]. Cloud computing is cost-efficient where users do not need to buy technical infrastructure, software, and information. In addition, cloud computing is distributed so it is more likely to face security threats and interaction as privacy violations, Denial of service, and unauthorized access [2].

One of defense lines for such environment is applying intrusion detection to protect cloud resources and services from intrusive activities [3].

Teodoro [4] has defined intrusion detection as "a security tool like other measures such as antivirus software and firewalls, are proposed for security, to become more powerful, of information and communication system". Therefore, Intrusion detection aims to control the events occurring on a network in an intelligent way.

James Anderson that focuses on how to make computers has published the first study of intrusion detection system in 1980 and systems secure. To accomplish this task, there was a need to use auditing files to detect unauthorized access [5].

Researchers in [6] announced that Intrusion detection system has different classification; according to the data source or the detection method.

Data source: which depends on the source of data. According to this factor IDS can be categorized into:

Host-based intrusion detection systems are aimed at collecting information about activity on a particular single system, or host, typically installed on a machine [5].

Network-based intrusion detection systems "These systems collect information from the network itself" [5].

Detection Method: According to this factor IDS can be categorized into Signature Based Intrusion Detection, also called misuse detect new attacks whose behavior unrecorded; this type only detects known attacks matched detection, records of known attacks are stored, and the behavior matched with the stored record. Intrusion depends on stored attack. The main disadvantage of this type concerns about mismatching with new, it cannot detect new attack only can record the behavior.

Anomaly Based Intrusion Detection System, describe a process of detecting abnormal activities on a network, beyond its gained knowledge. Researchers in [7] used predefined rules, classes and attributes identified from training data, set of classification rules, parameters and procedures inferred [8].

Anomaly Based Intrusion Detection System has two main advantages over Signature Based Intrusion Detection System: Firstly, Anomaly Based Intrusion Detection System can detect insider attacks or account theft very easily. If a real user or someone using a stolen account starts performing actions outside the normal user-profile, it generates an alarm. Because the system is based on customized profiles; it is very difficult for an attacker to know with certainty (what activity he can do without firing an alarm). However, the largest benefit of intrusive activity doesn't depend on specific traffic that represents known intrusive activity (as in Signature Based Intrusion Detection System). Secondly, Anomaly detection system can potentially detect an attack on the first time of use.

2. Related Work

Different researchers have used several Artificial Intelligent techniques for intrusion detection such as fuzzy logic, Neural Network and Genetic Algorithms for detecting network attacks, especially in a cloud environment which considered as an intended target for the intruders to exploit the weak points.

In [9], researchers have designed and implemented a Knowledge-Based Anomaly Intrusion Detection framework, which based on the Hyperbolic Hopfield Neural Network with anomaly, trained by using KDD'99 datasets, the Hyprolic Hopefield Neural Network is more efficient than Genetic Algorithms and Fuzzy Neural Network with a 95 % rate.

Wang in [10] proposed a new approach, called FC-ANN, based on ANN and fuzzy clustering, to give better detection rate and less false positive rate. This approach, firstly uses c-means fuzzy clustering technique. Each subset was assigned to separate ANN models, to generate different training subsets.

The different training subsets, based on different ANN models are trained to formulate different base models, at last they uses fuzzy aggregation employed to aggregate these results, the best average precision was 96.71 to report, when cluster number k= 6.

Selman in [11] has suggested an IDS model based on Fuzzy Logic. The model consists of three parts, Input Reduction System (IRS), which uses Principal Component analysis, it reduces the number of inputs from 41 to 13, Classification System, which uses Fuzzy C Means to create data clusters based on training data and Pattern Recognition System based on the Nearest Neighborhood method, which classifies new-coming data records to their expert clusters. Based on different attack types, the system performance in classification process is different and the best performance is achieved for PROBE attack, with a 99.3% success rate, and the best performance in pattern recognition is achieved for U2R with 58.8% of success rate.

Researchers in [12] have offered an intrusion detection approach, called FC-ANN, based on ANN and fuzzy clustering. In fuzzy clustering technique uses c-means, the heterogeneous training set is divided into several homogenous subsets. So involvement of each sub training set is reduced and then the detection performance is increased. In this way the system becomes more efficient, stable, and they efficiently stable the drawbacks –lower detection precision, weaker detection stability.

Lin and others in[13] have shown an evidence in the improvement for anomaly intrusion detection. First, it consists of feature

selection for anomaly intrusion detection. Second detect a new attack by the obtained decision rules from the dataset. Their proposed work tests anomaly intrusion detection over the KDD'99 dataset. Simulated annealing and Supported vector machine are implemented to find the best feature subsets. Simulated annealing and Decision Tree have proven their efficiency in generating decision rules to detect new attacks which gain 99.96% rate.

Shanmugavadivu and Nagarajan in [14] have developed an anomaly based intrusion detection system in detecting the intrusion behavior within a network. A fuzzy decision-making module was designed to build the system more accurate for attack detection, using the fuzzy inference approach. Firstly, the definite rules were generated by mining the single length frequent items from attack data as well as normal data. Then, fuzzy rules were identified, then these rules will be given to fuzzy system, which classify the test data, by analyzing the result, the overall performance of the proposed system is significantly improved and it achieves more than 90% accuracy for all types of attacks.

Panja and others in [15] have introduced the classification of more accurate way of classifying traffic network, also introduced the use of Genetic Algorithms (GA) with the adaptive NeuroFuzzy Inference System (ANFIS) to improve data classification and achieve better results. GA uses a set of genetic operators such as mutation, crossover and selection on current population to spawn similar patterns that will be used over and over until a particular criterion is met, at the first expert of the system; the detecting rate was 92.74%, on second 94.87%.

Khazaee and Faez in [16] have proposed a new approach for intrusion detection and traffic classification. They use hybrid fuzzy clustering and neural networks. This method has high performance in terms of precision, recall, f-value, Detection Rate, False Alarm Rate, accuracy, and ROC curve analysis than the other works. So, training dataset is divided into two groups: (1) suitable, wanted, dataset sample for training (2) Unsuitable, Unwanted, dataset which contains removed-Dataset. For training purposes, both groups will be used and the achieved accuracy was 99.5%, while the false alarm rate was 0.45 %.

3. KDD'99 Dataset

For conducting experiments on proposed work, there is a need to use the dataset benchmark for comparison methods performance through increasing Detection Rate and false alarm. KDD'99 [17] is the most appropriate dataset benchmark will be used in experiments.

Different researchers use KDD'99 to compare their results with others. The origin of KDD'99 was DARPA.

DARPA was prepared and managed by MIT Lincoln Labs. The objective was to survey and evaluate research in intrusion detection. A standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment, was provided. The KDD'99 intrusion detection contest uses a version of this dataset.

KDD Cup 99 analysis supplies useful information in the expansion of intrusion detection systems. The classification of records in the KDD dataset into normal &attack records involves mining rules involving the features present in the data set. The number of features in KDD huge, any IDS objective is to gain the most relevant set of reduced feature set, although there are a lot of algorithms for selecting the relevant features from the KDD dataset, so the fuzzy logic selects the reduced features (KDD Cup 99).

4. Proposed Work

Here, a new approach will be presented by adapting a hybrid model consists of Hopefield ANN and Simulating Annealling as an aggregator.The general framework for Anomaly IDS divided into several stages :

4.1 Dataset Grouping

KDD'99 dataset will be processed in parallel by assigning dataset into five groups according to the type of attack to be detected. These Groups denoted by Ki; where $i = DOS$, Probe, U2R, R2L and Normal.

4.2 Hopfield Artificial Neural Network (HANN)

In this stage parallel detecting will be performed by using Five HANNs. From previous stage, each one of these groups will be assigned to separate HANN. The result from every HANN will be progressed to the third stage.

The Hopfield artificial neural network was one of neural network types that had been investigated by John Hopfield in 1980s[18]. The Hopfield network has not specified number of inputs or output neurons. Also, all inputs and outputs are fully connected as in fully mesh network in both directions. HANN applied input simultaneously for all neurons and recuersively and continuously compute its output until reaching desired detection precious. Figure 1 presents an example on HANN.

4.3 Simulating Anealling aggregator

In this stage, all Hopfiled ANN's results will be taken as an input to simulating annealing [20] which assigned to each type of attack for providing final decisions whether the incoming request is attack or normal. Fig.2. represents proposed work stages.

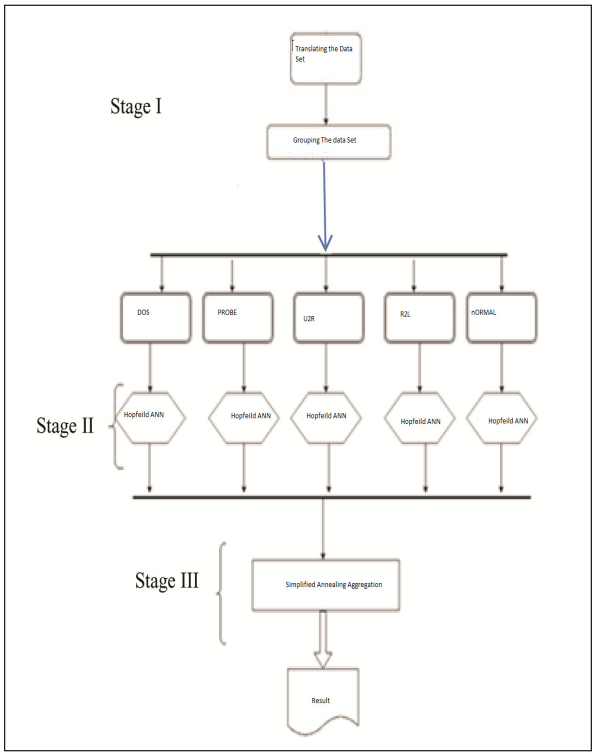

Figure 1. Detection Stages

5. Performance Measurements

In order to compare results and specify which achieve an enhancement, different measures have been used to judge on the performance of proposed methods and these measurements are True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) and Detection Rate (DR).

In [21,22] and others announced that intrusion detection evaluation problem and its solution effects on the choice of the suitable intrusion detection system for a particular environment depending on several factors. The most basic of these factors are the false alarm rate and the detection rate; they are calculated from the main four instances True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

In [21,22] and others used the following three equations to measure the standard in the proposed model:

Detection Rate (DR) can be defined as " the number of intrusion instances detected by the system (True Positive) divided by the total number of intrusion instances present in the test set". As shown in equation (1):

$$DR = P/(TP + FN) \qquad \text{..... (1)}$$

False alarm rate (FAR): can be defined as " the number of normal' patterns classified as attacks (False Positive) divided by the total number of normal' patterns ". As shown in equation (2):

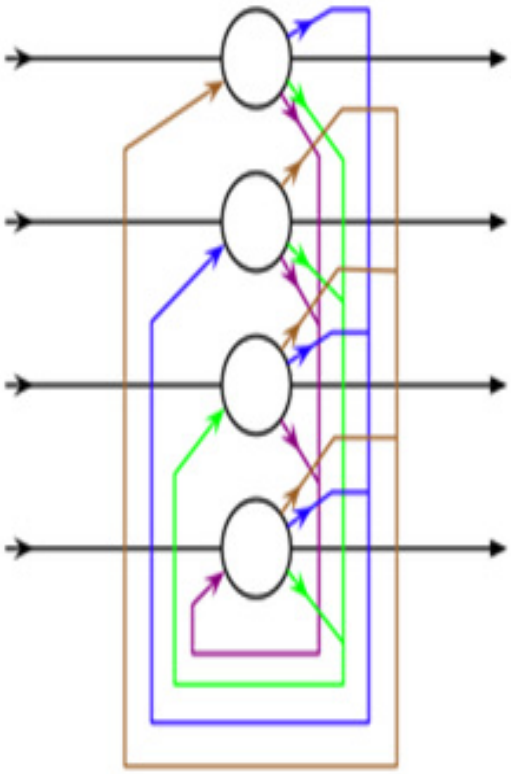

Figure 2. HANN with 4 neourns and 12 connections[19].

$$FAR = FP/FP + TN) \quad \text{…..……} \quad (2)$$

Accuracy (AC): can be defined as " the proportion of the total number of the correct predictions to the actual data set size". As shown in equation (3):

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad \text{…..……} \quad (3)$$

6. Results and Conclusion

This paper proposed a HANN and Simulating annealing that can be deployed to satisfy one of the security requirements of network environment especially cloud environment. According to the primary and earlier experiments, the proposed model achieved an <=93%, which can be considered as an accepted detection rate compared with methods in [9, 10, 13, 15] as shown in Table I. For gain better detection rate, more enhancement will be conducted by exploring the impact of network features in detection rate.

Method Rate	Detection
[9]	95%
[10]	96.71%
Proposed work	<=93%
[13]	99.96%
[15]	92.74%

Table 1. Detection Rate

References

[1] Kholidy Hisham, A., Baiardi Fabrizio., Hariri Salim. (2012). A Hierarchical Intrusion Detection System For Clouds: Design and Evaluation, *International Journal on Cloud Computing: Services and Architecture* (IJCCSA).

[2] Raghav Iti., Chhikara Shashi., Hasteer Nitasha. (2013). Intrusion Detection and Prevention in Cloud Environment: a Systematic Review, *International Journal of Computer Applications*.

[3] Gyanchandani Manasi., Rana J. L., Yadav R. N. (2012). Taxonomy of Anomly Based Intrusion Detection System: A Review, *International Journal of Scientific and Research Publications*.

[4] Teodoro, P., García., Verdejo, J., Díaz., Fernández, G., Maciá., Vázquez, E. (2009). Anomaly-basednetwork intrusion detection: *Techniques, systems and challenges*, Science Direct.

[5] SANS. (2001). The History and Evolution of Intrusion Detection, Institute Reading Room site.

[6] Poston Howard, E. (2012). A Brief Taxonomy of Intrusion Detection Strategies, Aerospace and Electronics Conference (NAECON), 2012 *Institute of Electrical and Electronics Engineers* (IEEE) National.

[7] Adaobi, O., Ghassemian, M. (2009). Analysis of an anomaly-based intrusion detection system for wireless sensor network, *International Conference onCommunications Engineering*.

[8] Kaur Harjinder., Gill Nivit. (2013). Performance Comparison of Host based and Network Based Anomaly Detection using Fuzzy Genetic Approach (FGA), *International Journal of Computer Trends and Technology* (IJCTT).

[9] Jabez, J., Muthukumar, B. (2007). Intrusion Detection System: Time Probability Method And Hyperbolic Hopfield Neural Network, *Journal of Theoretical and Applied Information Technology*.

[10] Wang Gang., Hao Jinxing., Ma jian., Huang Lihua. (2010). a new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering, *Expert Systems with Applications International Journal*.

[11] Selman Alma Husagic. (2011). Intrusion Detection System using Fuzzy Logic, *Southeast Europe Journal of Soft Computing*.

[12] Gaikwad, D. P., Jagtap Sonali., Thakare Kunal., Budhawant Vaishali. (2012). Anomaly Based Intrusion Detection System Using Artificial Neural Network and Fuzzy Clustering,International *Journal of Engineering Research& Technology* (IJERT).

[13] Lin Shih-Wei., Ying Kuo-Ching., Lee Chou-Yuan., Lee Zne-Jung. (2012). An intelligent algorithm with feature selection and decision rules applied to anomaly intrusion detection, *Expert Systems with Applications International Journal*.

[14] Shanmugavadivu, R., Nagarajan, N. (2013). Network Intrusion Detection System using Fuzzy Logic, *Indian Journal of Computer Science and Engineering* (IJCSE).

[15] Panja Biswajit., Ogunyanwo Olugbenga., Meharia Priyanka. (2014). Training of Intelligent Intrusion Detection System usingNeuro Fuzzy, *Institute of Electrical and Electronics Engineers* (IEEE).

[16] Khazaee Saeed., Faez Karim. (2014). A Novel Classification Method Using Hybridization of Fuzzy Clustering and Neural Networks for Intrusion Detection, *I.J.Modern Education and Computer Science*.

[17] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[18] LEE, B. W., Sheu, B J. (1991). Modified Hopfield Neural Network for Retrieving the Optimal Solution, *IEEE Transaction on Neural Networks*, 2 (1), January 1991

[19] http://en.wikipedia.org/wiki/Hopfield_network .

[20] Russell, S., Norvig, P. (2010). Artificial Intelligence: A Modern Approach, 3rd Edition, Prentice-Hall.

[21] Elhamahmy, M. E. N., Elmahdy Hesham., Saroit Imane, A. (2010). A New Approach for Evaluating Intrusion Detection System, *International Journal of Artificial Intelligent Systems and Machine Learning*.

[22] Chen Rung-Ching., Cheng Kai-Fan., Hsieh Chia-Fen. (2009). Using Rough Set And Support Vector Machine For Network Intrusion Detection, *International Journal of Network Security & Its Applications* (IJNSA).