Intrusion Detection Using Neural Network: A Literature Review

Asma Abbas Hassan[1], Alaa F. Sheta[2], Talaat M. Wahbi[3]

[1]Computer Science, Sudan University of Science & Technology Khartoum, Sudan

[2] Computer Science Department, Taif University, Taif , saudiarabia,

[3]Computer Science, Sudan University of Science & Technology Khartoum, Sudan

Abstract: ***Nowadays the computer security is important in our society,. Because of the wide use of computer networks and its application, it becomes imperative to detect the network attacks to protect the information security.therefor, anyone using a computer is at some risk of intrusion, even if he is not connected to the Internet or any other network . If the computer is left unattended, any person can attempt to access and misuse the system. The problem is, however, greater if the computer is connected to a network, especially the Internet. Any user from any place in the world can reach the computer remotely and may attempt to access private information. Solving the problem of attack detection using intrusion detection against computer networks is being a major problem in the area of network security. The intrusion detection system meets some challenges, and there are different approaches to deal with these challenges, neural network and machine learning is the best approaches to deal with it. In this paper we will illustrate different approaches of Intrusion detection system using neural network in briefly, and their advantages and disadvantages.***

Keywords: intrusion detection system, HIDS, NIDS, Hybrid IDS, anomaly detection, misuse detection

1. Introduction

All of our needs are becoming on the network, and the wide spread of computer network and its application lead to various types of attacks from intruders and hackers for different purposes. This may lead to the terrible disaster for the network users, that is why the network connection has to be very secure and protected every one's privacy. Doing this by computer manufactures in hardware will be very difficult from both technical and economical views. In this paper, we will discuss and analysis various research papers that used neural network to detect and prevent intrusion in meaning of developing the effective intrusion detection systems for computer systems and computer networks.

1.1 Intrusion detection definition

Intrusion detection is used to catch intrusions or attacks into computer and network systems when they want to violate the system security and privacy [1], as illustrated in figure (1) below:

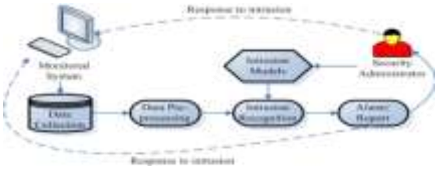

Figure 1: intrusion detection

1.2 Intrusion Detection Overview

Any intrusion detection system has four major categories that must be respected and put in mind when intrusion detection system mentioned.

These categories are:
1) Intrusion type
2) Detection behavior.
3) Detection approaches.
4) System types.

The categories are illustrated in the figure (2) below:[2]

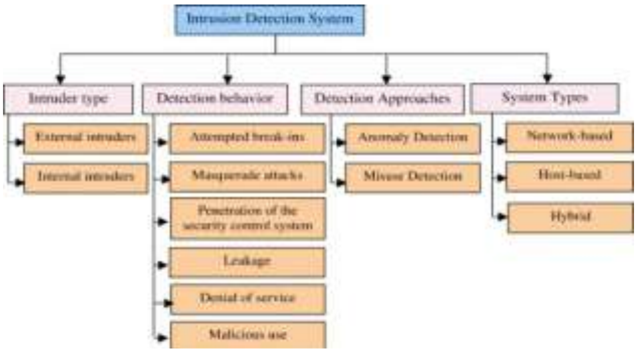

Figure 2: Intrusion detection system categories

1.2.1. Intruders Types
1) External intruders.
2) Internal intruders.

1.2.2. Detection Behaviour
1) Attembted break-ins
2) Masqurad attaccks
3) Penetration of the security control system
4) Leakage
5) Denial of serviece
6) Malicious use

1.2.3 detection approach

a) Anomaly detection

Anomaly detection is an approach used to identify unusual patterns that do not conform to expected behavior, called outliers. It has many applications in business, in intrusion detection (identifying strange patterns in network traffic that could signal a hack), fraud detection in credit card transactions and fault detection in operating environments.[17]

b) Misuse detection

Misuse detection is an approach to detecting computer attacks. In a misuse detection approach, abnormal system behaviour is defined first, and then all other behaviour is defined as normal. It stands against the anomaly detection approach which utilizes the reverse: defining normal system behaviour first and defining all other behaviour as abnormal. With misuse detection, anything not known is normal. An example of misuse detection is the use of attack signatures in an intrusion detection system. Misuse detection has also been used more generally to refer to all kinds of computer misuse[15]

1.2.4 System Types:

We can classify the Intrusions Detection into two main categories. They are as follows:

1) **Host Based Intrusion Detection Systems (HIDSs),** which evaluate information found on a single or multiple host systems, including contents of operating systems, system and application files, as illustrated in fig (3). [3]

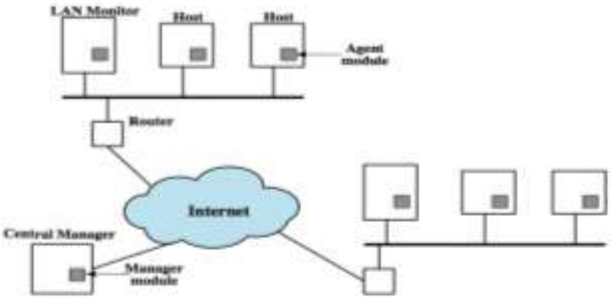

Figure 3: Host Based Intrusion Detection System

2) **Network Based Intrusion Detection Systems (NIDSs),** which evaluate information captured from network communications, analyzing the stream of packets which travel across the network, as illustrated in fig (4). [3]

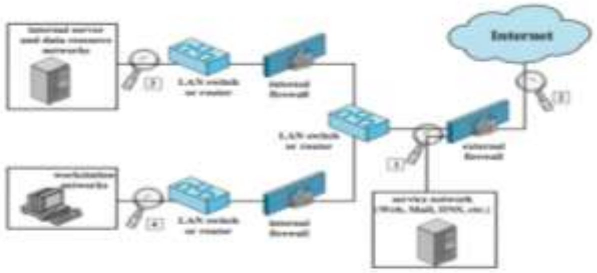

Figure 4: Network Based Intrusion Detection System

(3) Hybird Intrusion Detection Systems

Some of the most current intrusion detection system only uses one of the two detection methods, misused detection or anomaly detection both of them have their own limitations, this is the technique which combines misuse detection system and anomaly detection system is known as hybrid intrusion detection system or we can say that the technique [16] which combines the network intrusion detection system and host intrusion detection system is known as hybrid intrusion detection system.

1.3 Components of Intrusion Detection System

Any intrusion detection system must consist of three main components [4]

(1) The data source(event generator), it can be categorized into four categories, namely Host-based monitors, Network-based monitors, Application-based monitors and Target-based monitors.

(2) The analysis engine. Which takes information from the data source and examines the data for attacks. There are two ways of analysis attacks:
   - Misuse/Signature-Based Detection: This detects intrusion that follows well-known patterns of attacks (or signatures). [5,6]
   - Anomaly/Statistical Detection: An anomaly based detection engine will search for something rare or unusual. [7]
   - The response manager. Which works only when finding possible intrusion attacks on the system, as illustrated in fig (5)

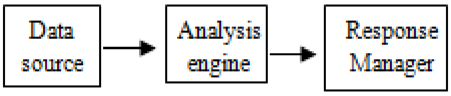

Figure 5: Components of Intrusion Detection System

2. Description of Dataset Taken for Experimentation

We take the KDD CUP 99 data set for testing and evaluating the proposed method. The United States Defense Advanced Research Projects Agency (DARPA), funded in 1998 an "Intrusion Detection Evaluation Program (IDEP)" administered by the Lincoln Laboratory at the Massachusetts Institute of Technology. This program built the KDD CUP data set to use it in any research in the field of intrusion detection. [8]

2.1 Records in KDD data set:

The records in KDD data set are divided into two main parts that is labeled and unlabeled records.

1) Each labeled record consisted of 41 attributes (features) and one target value. Target value is used to describe the attack category name. There are around 5 million (4,898,430) records in the labeled dataset.
2) Unlabeled dataset is used as test data. [8], table(1) illustrates.

Table 1: List of features in KDD cup 99 dataset

F1	*Feature name*	*Description*
1	Duration	Length (no. of seconds) of the connection
2	protocol_type	type of the protocol, say TCP, UDP, etc.
3	Service	network service on the destination, say http, telnet, etc.
4	flag	normal status or error status of connection
5	src_bytes	Number of data bytes from source to destination
6	dst_bytes	Number of data bytes from destination to source
7	Land	1 if connection is to/ from the same port / host; otherwise 0.
8	wrong_fragment	Number of ``wrong" fragments.
9	Urgent	Number of urgent packets.
10	Hot	Number of ``hot" indicators.
11	num_failed_logins	Number of failed login attempts.
12	logged_in	1 if successfully logged in; otherwise 0
13	num_compromised	Number of ``compromised" conditions
14	root_shell	otherwise 0 1 if root shell is obtained
15	su_attempted	otherwise 0. 1 if ``su root" command attempted
16	num_root.	Number of ``root" accesses.
17	num_file_creations	Number of file creation operations.
18	num_shells	Number of shell prompts.
19	num_access_files	Number of operations on access control files.
20	num_outbound_cmds	Number of outbound commands in an ftp session.
21	is_hot_login	1 if the login is fit in to the ``hot" list; otherwise 0.
22	is_guest_login	1 if the login is a ‘guest’; otherwise 0
23	Count	connection in the earlier period of two seconds. Number of connections to the same host as the current
24	srv_count	Number of connections to the same service as the current connection in the earlier period of two seconds
25	serror_rate	%of connections having ``SYN" errors
26	srv_serror_rate	% of connections having ``SYN" errors
27	rerror_	% of connections having ``REJ" errors
28	srv_rerror_rate	% of connections having ``REJ" errors
29	same_srv_rate	% of connections to the similar service
30	diff_srv_rate	% of connections to dissimilar services
31	srv_diff_host_rate	% of connections to dissimilar hosts
32	dst_host_count	count for target /destination host
33	dst_host_srv_count	srv_count for target /destination host
34	dst_host_same_srv_rate	same_srv_rate for target /destination host
35	dst_host_diff_srv_rate	diff_srv_rate for target /destination host
36	dst_host_same_src_port_rate	same_src_port_rate for target /destination host
37	dst_host_srv_diff_host_rate	diff_host_rate for target /destination host
38	dst_host_serror_rate	serror_rate for target /destination host
39	dst_host_srv_serror_rate	srv_serror_rate for target /destination host
40	dst_host_rerror_rate	rerror_rate for target /destination host
41	dst_host_srv_rerror_rate	srv_serror_rate for target /destination host

2.2 Features o2.2 Fe2

2.2.1 Features of KDD data set:

The features of KDD 99 data set can be classified into three groups:

2.2.2 Basic features:

All the attributes encapsulate in this category can be gained from a TCP/IP connection. It might cause a delay in detection. [9]

2.2.3 Traffic features:

The features in this category are computed with respect to a window interval and are divided into two groups, table(2) illustrates.

2.2.3.1“same host” features:

This group tests the connection in the past 2 seconds, which has the same target host as the current connection, and it calculates statistics related to protocol behavior and service.

2.2.3.2“same service” features:

This group tests the connection in the past 2 seconds, which has service as the current connection. [9]

Table 2: Traffic features computed using a two-second time window.

Feature Name	*Description*	*Type*
count	number of connections to the same host as the current connection in the past two seconds	continuous
	Note: The following features refer to these same-host connections.	
serror_rate	% of connections that have ``SYN" errors	continuous
rerror_rate	% of connections that have ``REJ" errors	continuous
same_srv_rate	% of connections to the same service	continuous
diff_srv_rate	% of connections to different services	continuous
srv_count	number of connections to the same	continuous

	service as the current connection in the past two seconds	
	Note: The following features refer to these same-service connections.	
srv_serror_rate	% of connections that have ``SYN" errors	continuous
srv_rerror_rate	% of connections that have ``REJ" errors	continuous
srv_diff_host_rate	% of connections to different hosts	continuous

2.2.3 Content features

This category used when detecting certain type of attacks that need some features to be able to look for suspicious behavior in the data portion, e.g., number of failed login attempts. [9], table(3) illustrates.

Table 3: Content features within a connection suggested by domain knowledge

feature name	*description*	*type*
hot	number of ``hot" indicators	continuous
num_failed_logins	number of failed login attempts	continuous
logged_in	1 if successfully logged in; 0 otherwise	discrete
num_compromised	number of ``compromised" conditions	continuous
root_shell	1 if root shell is obtained; 0 otherwise	discrete
su_attempted	1 if ``su root" command attempted; 0 otherwise	discrete
num_root	number of ``root" accesses	continuous
num_file_creations	number of file creation operations	continuous
num_shells	number of shell prompts	continuous
num_access_files	number of operations on access control files	continuous
num_outbound_cmds	number of outbound commands in an ftp session	continuous
is_hot_login	1 if the login belongs to the ``hot" list; 0 otherwise	discrete
is_guest_login	1 if the login is a ``guest"login; 0 otherwise	discrete

3. Artificial Neural Networks

An Artificial Neural Network consists of a collection of input elements that when processed with respect to a hidden element, get the desired output elements. The result of the processing which we call the output is determined by the characteristics of the elements and the weights associated with the interconnections among them. The network gets the desired output by modifying the connections between the nodes [10,11].

A neural network trained to recognize the characteristics of the matched data that has been analyzed, and give the probability estimation, which can be very accurate (100%), this accuracy depends on the system training times. The Neural Network system is trained and been refined till the accuracy is being as best as possible and reach satisfied level.

3.1 Neural Network Intrusion Detection Systems

A limited amount of research has been made by the neural network application in detecting computer intrusions. Artificial Neural Networks are the most suitable approach to detect computer intrusions, it can solve major problems that other current approaches cannot deal with[11]. Neural networks were mainly proposed to recognize the system users' characteristics and determine statistically considerable differences from the user's behavior.

First must collect the data representing normal and abnormal behavior to train the Neural Network. After training is done, then we will get a certain number of performance tests with real network traffic and attacks.

Artificial Neural Networks have also been proposed for use in the detection of computer viruses. In [12] and [13] Neural Networks were proposed as statistical analysis approaches in the detection of viruses in computer networks.

The Neural Network model in [13] was used a single layer of Neurons to represent knowledge from a particular domain in the form of a geometrically organized feature map. The model was made to learn the behavior of the normal system activity and then get the differences from the norm that may be an indication of a virus.

3.2 Advantages of Neural Network-based Intrusion Detection Systems

The flexibility is the most important advantage of a neural network in the detection. That the Neural Network can analyze the data from the network, even if the data is incomplete or distorted. The Neural Networks has other advantages that are the speed. Because the output of a Neural Network is expressed in the form of a probability, the Neural Network provides a predictive capability for the detection of any attack.

Another important advantage of Neural Networks is the ability of the Neural Network to "learn" the characteristics of any attacks and classify it according to attack types.

3.3 Disadvantages of Neural Network-based Intrusion Detection Systems

The Neural Networks have many limitations that lead to the rare usage in detection, one of that is the training requirements of the Neural Network. Because of the complexity of the training method that were used, and need a very large amount of data and multi refinements to get accurate results. The training routine requires a very large amount of data to ensure that the results are statistically accurate. However, the most significant and recent disadvantage of applying Neural Networks to intrusion detection is the "black box". The "Black Box Problem" now is an on-going area of Neural Network research [14].

4. Future Work

An efficient Intrusion Detection Systems can be constructed with flexible, high learning rate, accuracy adaptability,and effective results with very low error rate, by using Artificial Neural Networks approaches.

5. Conclusion

This paper presented an overview of Intrusion Detection System using Artificial Neural Network technologies. And compared the different Artificial Neural Network technologies for Intrusion Detection and their advantages and disadvantages. Finally, a discussion of the future ANN technologies, which can be used to help with network security and to enhance the ability of computer systems to detect intrusions.

References

[1] Data Mining With Application to Fraud and Intrusion Detection: Results from the JAM Project by Salvatore J. Stolfo, Wei Fan, Wenke Lee, Andreas Prodromidis, and Philip K. Chan.

[2] A Detailed Analysis of the KDD CUP 99 Data Set Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani

[3] J. P. Planquart, “Application of Neural Networks to Intrusion Detection”, SANS Institute Reading Room.

[4] R. G. Bace, “Intrusion Detection”, Macmillan Technical Publishing. 2000.

[5] S. Kumar, E. Spafford, “A Software architecture to Support Misuse Intrusion Detection” in The 18th National Information Security Conference, pp. 194-204. 1995.

[6] K. Ilgun, R. Kemmerer, P. A. Porras, “State Transition Analysis: A Rule-Based Intrusion Detection Approach”, IEEE Transaction on Software Engineering, 21(3):pp. 181-199. 1995.

[7] S. Kumar, “Classification and Detection of Computer Intrusions”, Purdue University, 1995.

[8] A Detailed Analysis of the KDD CUP 99 Data Set Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani .

[9] Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context , Maheshkumar SabhnaniEECS Dept, University of Toledo Toledo, Ohio 43606 USA and Gursel SerpenEECS Dept, University of Toledo Toledo, Ohio 43606 USA

[10] Fox, Kevin L., Henning, Rhonda R., and Reed, Jonathan H. (1990). A Neural Network Approach Towards Intrusion Detection. In Proceedings of the 13th National Computer Security Conference.

[11] Hammerstrom, Dan. (June, 1993). Neural Networks At Work. IEEE Spectrum. pp. 26-53.

[12] Denault, M., Gritzalis, D., Karagiannis, D., and Spirakis, P. (1994). Intrusion Detection: Approach and Performance Issues of the SECURENET System. In Computers and Security Vol.13, No. 6, pp. 495-507

[13] Gang Wang, Jinxing Hao, Jian Ma and Lihua Huang (2010), "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering", Expert Systems with Applications, pp. 1-8.

[14] Fu, L. (1992). A Neural Network Model for Learning Rule-Based Systems. In Proceedings of the International Joint Conference on Neural Networks. pp. (I) 343-348.

[15] Helman, Paul, Liepins, Gunar, and Richards, Wynette, "Foundations of Intrusion Detection," The IEEE Computer Security Foundations Workshop V, 1992

[16] Harley Kozushko “Intrusion Detection: Host-Based andNetwork-Based Intrusion Detection Systems” Duanyang Zhao, Qing XiangXu, Zhilinfeng“analysis and design for intrusion detection system based on data mining”978-0-7695-3987-4/10$26.002010IEE

[17] ABE, N., Zadrozny, B., and Langford, J. 2006. Outlier detection by active learning. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM Press, New York, 504–509