computing@tanet.edu.te.ua
www.tanet.edu.te.ua/computing

ISSN 1727-6209
International Scientific
Journal of Computing

INTRUSION RECOGNITION USING NEURAL NETWORKS

Vladimir Golovko[1], Pavel Kochurko[2]

[1] Brest State Technical University, Moskovskaja str. 267, 224017 Brest, Belarus, gva@bstu.by
[2] Brest State Technical University, Moskovskaja str. 267, 224017 Brest, Belarus, paulermo@tut.by

Abstract: *Intrusion detection techniques are of great importance for computer network protecting because of increasing the number of remote attack using TCP/IP protocols. There exist a number of intrusion detection systems, which are based on different approaches for anomalous behavior detection. This paper focuses on applying neural networks for attack recognition. It is based on multilayer perceptron. The 1999 KDD Cup data set is used for training and testing neural networks. The results of experiments are discussed in the paper.*

Keywords: *Neural networks, intrusion detection systems, network attacks, attack recognition.*

1. INTRODUCTION

The rapid and extensive growth of Internet technology increases the importance of protecting computer networks from attacks. In the last years the number of network attacks has been raised very promptly that has led to significant problems in different companies. For instance some companies like Yahoo were attacked by DoS (denial of service), costing them millions of dollars.

Intrusion detection systems (IDS) are used as a computer network security tool and permit to alert an administrator in case of attack. The main goal of IDS is to detect and recognize network attacks in real time. Nowadays there exist different approaches for intrusion detection. It is signature analysis, rule-based method, embedded sensors, neural networks, artificial immune systems [1, 2, 3, 4, 5, 6] and so on. The most of these IDS can detect the known attacks and have poor ability to detect new attacks.

In last years a neural network techniques have been applied and investigated for intrusion detection [7, 8, 9, 10] Such approaches are based on different strategies. So, one of them for anomaly detection use analysis of the audit records, produced by the operating system [8]. The other one is based on network protocol analysis [9].

Among the most wide-spread neural networks are feedforward networks, namely multilayer perceptron (MLP). This network type has been proven to be universal function approximator [11]. Another important feature of MLP is the ability to generalization. Therefore MLP can be powerful tool for design of intrusion detection systems.

This paper presents applying of neural networks for intrusion detection through an examination of network traffic data. It has been shown that denial of service and other network-based attacks are presented in the network traffic data. Therefore using neural networks permits to extract nonlinear relationships between variables from network traffic and to design real-time intrusion detection systems.

We describe the intrusion recognition system, which is based on MLP.

The rest of the paper is organized as follows. The section 2 describes attack classification and training data set. In the section 3 the intrusion detection system is described, based on neural network approach. Section 4 presents experimental results. Conclusion is given in section 5.

2. ATTACK CLASSIFICATION AND KDD DATA SET

An event is a minimal unit with which modern protection tools operate. As soon as event breaks a policy of security, it at once is considered as a part of attack. Action or sequence of the connected actions of the intruder resulting in realization of threat by use of vulnerabilities is called attack to information system.

There are various types of attack classifications. For example, division into passive and active, external and internal attacks, deliberate and unintentional. It should be mentioned that many models of attacks are currently well known: ”one–to–one” or ”one–to–many”, i.e. attack proceeds from one point; ”many–to–one” and ”many–to–many”, i.e.

distributed or coordinated attacks; hybrid attacks also named the blended threat [12].

In the 1998 DARPA intrusion detection evaluation program, an environment was set up to acquire raw TCP/IP dump data for a network by simulating a typical U.S. Air Force LAN. The LAN was operated like a true environment, but being blasted with multiple attacks [13]. In 1999 sample data set of network traffic was presented at KDD'99 conference [14].

Attacks can be classified on the purposes of intrusion. Some of this categories were used in KDD data set [12, 14]:

Remote penetration, R2L – attacks which allow to realize the remote control of a computer through a network: unauthorized access from a remote machine.

Local penetration, U2R – the attack resulting in assigning of non-authorized access to the site on which it is started, unauthorized access to local superuser (root) privileges.

Remote denial of service, DoS – attack which allows to break functioning of system or to overload a computer through Internet.

Local denial of service, DoS – the attack, allowing to break functioning system or to overload a computer on which it is realized. An example of such attack is the hostile applet which loads the central processor an infinite cycle that results in impossibility of transaction processing of other applications.

Scanners, probing – analysis of the topology of a network, services accessible to attack, carrying out search of vulnerabilities on network hosts.

Sniffers – programs which "listen" to the network traffic. Using these programs it is possible to search automatically for identifiers and passwords of users, the information on credit cards, etc.

KDD database consists of 4940210 records where every record describes one TCP/IP connection. Only 20% of records represent normal connections. A connection is by a sequence of TCP packets during a duration whose starting time and ending time are both well defined, and data flow during this duration from a source IP address to a target IP address under some well defined protocol. Each connection is labeled as either normal or attack. In the latter case, the connection should be with exactly one specific attack type.

For each TCP/IP connection, 41 various quantitative and qualitative features were extracted [14]. This features can be divided into three categories: intrinsic features, i. e., general information related to the connection; traffic features, i. e., statistics related to past connections similar to the current one e. g., number of connections with the same destination host or connections related to the same service in a given time window or within a predefined number of past connections; content features, i. e., features containing information about the data content of packets that could be relevant to discover an intrusion [15]. Each connection record consists of approximately 100 bytes.

3. SYSTEM DESCRIPTION

Two approaches to intrusion detection are currently used. The first one, called misuse detection is based on the knowledge of attacker behavior. Intrusion detection system compares current network activity with the known patterns of behaviors of attackers attempting to penetrate the system. The second one, called anomaly detection involves identifying activities that vary from established behavior of users, or groups of users. Anomaly detection though is often highly difficult, as it must be tailored system to system, and sometimes even user to user, as behavior patterns and system usage can vary widely [12, 9, 16] .

Let's examine the block-diagram of the intrusion detection system (Fig. 1). It consists of several stages. At the beginning the system reads traffic data and sends it to the preprocessing module. The task of preprocessing module is to collect necessary data for neural networks from network traffic.

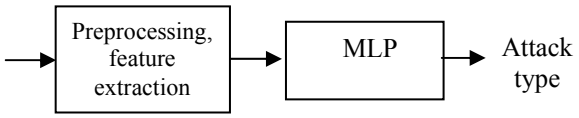

Fig. 1 - Block diagram of the network traffic processing Simplified IDS structure

Our intrusion detection system uses its own sniffer based on WinPCap driver to collect raw traffic data [17]. WinPCap is Windows port of UNIX pcap and is used for sniffing of network devices. It provides gathering data from IP, TCP, UDP, ICMP protocols. Every packet we receive from WinPCap has its header and body. The system analyzes the header data and calculates the parameters of TCP-connections. Every incoming and outgoing packet is analyzed and its parameters are added to the connection parameters. The following set of parameters of TCP-connections (Table 1) are selected by preprocessing module for training and testing of neural networks, like it is shown in Tables 3–5.

Such a system permits to identify and recognize the network attacks.

Let's consider the neural network for recognition of attack. This network is multilayer perceptron with 6 input units, 40 hidden units and 23 output units, where the number of the unit with maximal value

shows the type of recognized attack (Fig. 2). It should be noted that one MLP for each service has been used. The backpropagation algorithm is used for MLP training.

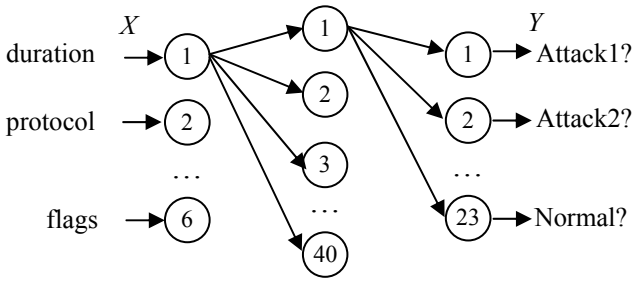

Fig. 2 – MLP structure

The results of experiments are discussed in the next section

4. EXPERIMENTAL RESULTS

To assess the effectiveness of the proposed intrusion detection approach, the experiments were conducted on the KDD Cup network intrusion detection data set [14]. Training data sets for recognition of attack consist of normal samples and attacks (Table 2) for each service.

Let's examine the recognition of attack types (Table 3). Table 4 shows the statistic of recognition attacks depending on attack type for some services. Table 5 shows the common results concerning attack recognition and detection for four categories. As can be seen MLP network can recognize 94,49% attack.

Table 1. Selected network traffic elements

Feature name	Description	Type
duration	length (number of seconds) of the connection	cont.
protocol type	type of the protocol, e.g. tcp, udp, etc.	discr.
service	network service on the destination, e.g., http, telnet, etc.	discr.
src bytes	number of data bytes from source to destination	cont.
dst bytes	number of data bytes from destination to source	cont.
logged in	1 if successfully logged in; 0 otherwise	discr.
flags	TCP/IP network flags	discr.

Table 2. Training data sets

Attack Type	# of normal samples	Total samples
auth	220	328
bgp	0	106
domain	3	116
eco i	109	207
finger	468	670
ftp	190	407
ftp data	350	457
http	219	442
pop 3	79	202
private	180	458
smtp	79	99
telnet	219	513

Table 3. Identification and recognition statistics depending on service

Service	True Alarms	False Alarms	Recogn. Correctly
1 auth	108 100%	0	108 100%
2 bgp	106 100%	0	0 0%
3 courier	108 100%	0	88 81,48%
4 csnet ns	126 100%	0	100 79,37%
5 ctf	97 100%	0	78 80,41%
6 daytime	103 100%	0	102 99,03%
7 discard	116 100%	0	89 76,72%
7 domain	113 100%	0	112 99,12%
8 domainu	0 0%	0	0 0%
9 echo	112 100%	0	89 79,46%
10 eco i	1253 100%	0	1149 91,7%
11 ecr i	281049 99,99%	0	280790 99,90%
12 efs	103 100%	0	79 76,7%
13 exec	99 100%	0	99 100%
14 finger	200 99,01%	3 0,64%	180 90%
15 ftp	414 97,41%	3 0,8%	409 98,79%

Service	True Alarms	False Alarms	Recogn. Correctly
18 host-names	104 100%	0	86 82,69%
19 http	2364 98,21%	220 0,36%	2362 99,92%
20 http 443	99 100%	0	81 81,82%
21 imap4	116 99,15%	0	82 70,69%
22 irc	1 100%	31 73,81%	1 100%
23 iso tsap	115 100%	0	96 83,48%
24 klogin	106 100%	0	82 77,36%
25 kshell	98 100%	0	82 83,67%
28 login	102 98,08%	0	102 100%
29 mtp	107 100%	0	83 77,57%
30 name	98 100%	0	78 79,59%
31 netbios dgm	99 100%	0	0 0%
32 netbios ns	102 100%	0	82 80,39%
33 netbios ssn	107 100%	0	0 0%
34 netstat	95 100%	0	1 1,05%
35 nnsp	105 100%	0	86 81,9%
36 nntp	108 100%	0	106 98,15%
37 other	1602 99,81%	93 1,65%	1228 76,65%
38 pop_2	101 100%	0	82 81,19%
39 pop_3	122 99,19%	0	119 97,54%
40 printer	109 100%	0	107 98,17%
41 private	103500 99,97%	2 0,03%	83900 81,01%
42 remote job	120 100%	0	101 84,17%
43 rje	111 100%	0	83 74,77%
44 shell	111 100%	0	111 100%
45 smtp	122 97,6%	28 0,29%	120 98,36%
46 sql_net	110 100%	0	0 0%
47 ssh	104 100%	0	102 98,08%
48 sunrpc	107 100%	0	86 80,37%
49 supdup	105 100%	0	77 73,33%
50 systat	115 100%	0	92 80%
51 telnet	250 85,03%	3 1,37%	246 98,4%
52 tftp_u	0	1 100%	0
53 time	103 100%	2 3,85%	103 100%
54 uucp	106 100%	0	80 75,47%
57 whois	110 100%	0	90 81,82%
58 X11	2 100%	8 88,89%	2 100%
59 Z39_50	92 100%	0	0 0%

Table 4. Statistics depending on attack types

Attack	Count	Detected	Recogn.
1 back	2203	2192 99,5%	2192 100%
2 buffer overflow	30	0 0%	0 0%
3 ftp write	8	2 25%	2 100%
4 guess passwd	53	49 92,45%	49 100%
5 imap	12	11 91,67%	1 9,09%
6 ipsweep	1247	1236 99,12%	1161 93,93%
7 land	21	21 100%	0 0%
8 loadmod-ule	9	0 0%	0 0%
9 multihop	7	1 14,29%	0 0%
10 neptune	107201	107177 99,98%	86445 80,6%
11 nmap	231	205 88,74%	99 48,29%
12 perl	3	0 0%	0 0%
13 phf	4	2 50%	2 100%
14 pod	264	259 98,11%	0 0%

Attack	Count	Detected	Recogn.
15 portsweep	1040	1038 99,81%	498 47,98%
16 rootkit	10	2 20%	2 100%
17 satan	1589	1578 99,31%	1522 96,45%
18 smurf	280790	280790 100%	280790 100%
19 spy	2	0 0%	0 0%
20 teardrop	979	977 99,8%	977 100%
21 warez-client	1020	427 41,86%	427 100%
22 warez-master	20	17 85%	16 94,12%

Table 5. Identification and recognition statistics depending on attack category

Category	Count	Detected	Recognized
1 dos	391458	391416 99,98%	370404 94,62%
2 u2r	52	2 3,84%	0 0%
3 r2l	1126	509 45,2%	497 97,64%
4 probe	4107	4057 98,78%	3280 79,86%

5. CONCLUSION

This paper describes the applying of MLP for attack recognition. In comparison with other approaches the neural networks permit to design the intrusion detection systems, which have ability to training and working in real time. The experiments have shown the efficiency of neural networks techniques.

6. ACKNOWLEDGMENT

This research is supported by the Grant of Belarus National Academy of sciences.

7. REFERENCES

[1]. M. Bishop, S. Cheung, C. Wee, J. Frank, J. Hoagland and S. Samorodin. *The treat from the Net*, IEEE Spectrum, 34(8), pp. 56-53, 1993.

[2]. D. Anderson, T. Frivold & A. Valdes. *Next-generation Intrusion Detection Expert Systems (NIDES): A Summary*, SRI International Technical Report SRI-CSL-95-07, 1995.

[3]. E. Spafford and D. Zamboni. *Data collection mechanisms for intrusion detection systems*, CERIAS Technical Report 2000-08, CERIAS, Pardue University, 1315 Recitation Building, West Lafayette, IN, 2000.

[4]. H. Debar, M. Becke & D.Simboni. *A Neural Network Component for an Intrusion Detection System*, In proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy, (1992).

[5]. C. Jirapummin and N. Wattanapongsakorn. *Visual Intrusion Detection using Self-Organizing Maps*, Proc. of Electrical and Electronic Conference (EECON-24), Thailand, Vol. 2, pp. 1343-1349, 2001.

[6]. S.C. Lee and D.V. Heinbuch. *Training a Neural Network Based Intrusion Detector to Recognize Novel Attacks*, Information Assistance and Security, pp. 40-46, 2000.

[7]. S. C. Lee, D. V. Heinbuch, *Training a Neural-Network Based Intrusion Detector to Recognize Novel Attacks*, IEEE Trans. on Systems, Man, and Cybernetics, Part A, 31, 2001, pp. 294-299.

[8]. A. K. Ghosh and A. Schwartzbard, *A Study in Using Neural Networks for Anomaly and Misuse Detection*, In Proc. of the USENIX Security Symposium, August 23-26, 1999, Washington, USA.

[9]. J. Cannady, *An adaptive neural network approach to intrusion detection and response*, PhD Thesis, School of Comp. and Inf. Sci., Nova Southeastern University,

[10]. J. M. Bonifacio et al., *Neural Networks applied in intrusion detection systems*, Proc. of the IEEE World congress on Comp. Intell. (WCCI'98), 1998.

[11]. K. Hornik, M. Stinchcombe, H. White, *Multy-layer feedforward networks are universal approximators*, Neural Networks, 2 pp. 359-366, 1989.

[12]. A. Lukatsky. Intrusion detection. Saint-Petersburg: BHV-Peterburg, 2003.

[13]. *MIT Lincoln Laboratory -DARPA Intrusion Detection Evaluation Web Page Template*. Training Data Attack Descriptions http://www.ll.mit.edu/IST/ideval/docs/1998/attacks.html

[14]. *1999 KDD Cup Competition.* http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[15]. W. Lee and S. J. Stolfo. *A framework for constructing features and models for intrusion detection systems*, ACM Trans. on Inform. and System Security, 3(4), 200, 227-261.

[16]. J. Cannady, *Applying Neural Networks to Misuse Detection*, In Proceedings of the 21st National Information Systems Security Conference.

[17]. *WinPcap: the Free Packet Capture Architecture for Windows*, NetGroup, Politecnico di Torino (Italy), http://winpcap.polito.it, 1999-2004.

Prof. Vladimir Golovko *was born in Belarus in 1960. He received M.E. degree in Computer Engineering in 1984 from the Moscow Bauman State Technical University. In 1990 he received PhD degree from the Belarus State University and in 2003 he received doctor science degree in Computer Science from the United Institute of Informatics problems national Academy of Sciences (Belarus). At present he works as a head of Intelligence Information Technologies Department and Laboratory of Artificial Neural Networks of the Brest State Technical University. His research interests include Artificial Intelligence, neural networks, autonomous learning robot, signal processing, chaotic processes, intrusion and epilepsy detection. He has published more than 150 scientific papers, including 3 books and 2 chapters of books.*

Pavel Kochurko, *2nd year PhD student, Lab. Of Artificial Neural Networks, Dep. Of Intelligent Informational Technologies, Brest State Technical University, Belarus. His research interests include computer network security, data-mining, artificial intelligence and neural networks. In 2002 he has won the prize of "First International Delphic Games of CIS-states" in web-design.*