Network Intrusion Detection Using Genetic Algorithm to find Best DNA Signature

THAER AL_IBAISI, ABD EL-LATIF ABU-DALHOUM, MOHAMMED AL-RAWI, MANUEL ALFONSECA *, ALFONSO ORTEGA*
King Abdullah school For Information Technology, University of Jordan
Escuela Politécnica Superior, Department of Computer Engineering, Universid Autonoma de Madrid, Madrid 28049, Spain*
ta0002@gmail.com ; a.latif@ju.edu.jo ; rawi@ju.edu.jo; manuel.alfonseca@uam.es; alfonso.ortega@uam.es

Abstract: - Bioinformatics is part of computer science that joins between computer programming and molecular biology. DNA consists of long sequence of nucleotides which formulates the genome. Our method is to generate normal signature sequence and alignment threshold value from processing the system training data and encode observed network connection into corresponding DNA nucleotides sequence, then to align the signature sequence with observed sequence to find similarity degree value and decide whether the connection is attack or normal. Number of DNA sequences makes up each population, and then new generations are produced to select the Signature with best alignment value with normal network connection sequences. This paper ends up with accuracy value and threshold score for detecting the network anomalies that no known conditions exist for them to be discovered in addition for percentage of generating false positive and true negative alarms.

Key-Words: - Sequence alignment, Intrusion detection system, Sequence encoding, Alignment threshold.

1 Introduction

In computer networks, computers are exposed to be attacked by external intruder. Many Computer networks have been attacked against their secure information, which in most cases result in financial loss crimes in addition to loss of valuable information. Intrusion can utilize the resources of the network in dangerous way [13]. Analysis of network flow contents is heavily helpful in searching for known attack patterns. Bioinformatics alignment methods help in handling the search for similarity areas in sequences' strings [4]; some of these methods can be useful in detection mechanism for network intrusions [14]. Our system needs to depend on predicting the safety of the network activities based on comparing the alignment score with pre-calculated threshold value that is generated from training data manipulation [2]. Training data is collecting from audit log of the network activities with labeled connection type; the label is either "normal" or "attack type", so every activity is entered to translated phase which convert the main command and its parameters into sequence of nucleotides then into Amino acid sequence which is saved beside each network connection request in the files. Testing data are records selected randomly from dataset and aligned to existing data so as to calculate the average score of the alignment against training data. The problem in finding all possible alignments is intractable [12]. Dynamic programming algorithm is so used to break down the problem into sub problems then find their partial solutions and generate final result [11]. Using global alignment method seems useful in detecting the intrusion attacks. We need to predict the safety of new and unknown network activities that we don't have exact match sequence for it.

2 Problem Formulation

Research for new intrusion detection techniques is very important, since in every day, new challenge for security breaches appear for IT people and make it hard to protect their networks or even discover the intrusions into their networks [7]. Researchers still try to reach a satisfying percentage of intrusion detection accuracy in case of false positive, false negative and hit ratio [14]. In addition to that, we need to have an intelligent technique and algorithmic way to decide on the network connections if it is dangerous or not [8].

3 Problem Solution

We use global sequence alignment from Bioinformatics to achieve better percentages for intrusion hit ratio. Global alignment is a similarity search method which takes two DNA sequences as input and produces a value as score for their similarities [8]. The higher the score between any two sequences the more similarities they have.

3.1 KDD Data Set

We use KDD'99 [1] data set as simulated audit data from computer network. In KDD'99 dataset files, each network activity is labeled with its type either attack or normal. Getting audit data for any network is not an easy job since it is forbidden to be distributed for security issues, so Lincoln Labs made it easier by simulating a typical U.S. Air Force LAN. They set up similar environment and enable it to acquire 9 weeks of raw TCP dump data. Also they simulate multiple attacks on that network to be the same as true Air Force LAN. The raw training data was about four gigabytes of compressed binary TCP dump data from seven weeks of network traffic. This was processed into about five million connection records. Similarly, the two weeks of test data yielded around two million connection records. Each connection is labeled as either normal, or as an attack, with exactly one specific attack type. Each connection record consists of about 100 bytes.

3.1.1 Attack Categories

Attacks can be categorized as the following:

Table 1 *Attack categories and their related attacks*

Category	Attacks
Denial Of Service – DoS	back, Neptune, pod, smurf, teardrop, process table, warezmaster, apache2, mail bomb.
Unauthorized To Root – U2R	http tunnel, ftp_write, multihop, buffer overflow, root kit, xterm, and ps.
Remote To Local – R2L	guess_passwd, named, snmpgetattack, xlock, send mail
Probe	ipsweep, nmap, port sweep, satan, mscan, saint

3.2 Intrusion Detection Methodology

For each record, the network activity is encoded into DNA sequence using certain encoding methodology. Normal behavior helps in differentiate between attack and normal activities. Normal behavior consists of threshold value and normal signature sequence. Threshold value is calculated from taking average of the alignment scores between normal signature sequence and normal sequences of sample activities that are randomly selected from training data set. Based on our system, the intrusion detection methodology implements the following steps:

1. Encoding audit data into DNA sequences.
2. Generate normal signature sequence and threshold value for system normal behavior.
3. Observed activity is encoded into DNA sequence and globally aligned with normal signature sequence.
4. Alignment score is compared with system threshold value to label the observed activity either attack or normal.

5. If the alignment score is greater than threshold value then network activity is normal, otherwise it is an attack and system needs to raise alert.

Normal DNA sequence signature is built from processing the encoded nucleotide sequences of the normal samples consist of network activities only. Normal signature sequence represents the normal behavior of the detection system on which we depend on it to differentiate between attack and normal behavior of the suspected network activities.

The challenge of building the signature is to find a suitable way of building a sequence that can represent as much as it can have from the sequences of the normal activities in the audited file, so we can guess that normal signature sequence may be different from audit data set to another one, and from system to system but the signature building algorithm is same for all. In the same time we need to keep its size suitable and fits with all the operations of alignment through the system. Although the longer signature supposed to contain many different shapes from the normal behavior, but the longer the signature sequence, the bad performance we get when repeatedly processing it, in addition to more space to store it.

Many methods to build normal signature sequence can be recommended but the most needed requirements to have from the output of these methods are shorter signature and rich containment it has for normal behavior shapes of the system data [10]. One of the methods to build the normal signature is to append the sequences of the encoded normal activities in the system. Although we contain all the normal behavior in the signature from previous method, but we end up with huge signature sequence to save and bad performance when align with it.

3.3 DNA Sequence Encoding Methodology

Our study needs to generate DNA nucleotide sequence for each network activity so as to input it in the sequence alignment process. The challenge is to encode different network audit parameter types and values into DNA nucleotides with the following:

a. Less space for saving them.

b. To have available combination of nucleotides for any new parameter values.

Since we need to try our method of encoding to see if it can benefit the process from capacity and performance saving.

We define mapping rules for our encoding methodology to be as the following:

1. Groups Identifiers: To increase the possibilities of finding more encoding sequence representations for different network parameters' values, we try to define a set of nucleotides that represent a fixed prefix for certain group of network activity parameters. The following are the group prefixes:

- AAA: This prefix is part from all Flags sequences.
- TTT: This prefix is part from all Protocols sequences.
- CCC: This prefix is part from all Services sequences.

We could run out of the codons and using all of them since some groups have more than 64 types (such as Services group), so we start duplicating the nucleotide sequence part that comes after prefix sequence as the following example:

HTTP => CCCTGC

To generate new sequence for NNSP from existing HTTP sequence:

NNSP => CCCTGC**TGC**

2. Static Parameters: There are known parameters that we can bound their expected values. These parameters are pre-defined and DNA nucleotides are defined for each value as the following:

   Flags:

   Table 2 *Network activity flags and their corresponding nucleotide sequence*

Flag	Nucleotides
SF	AAAGCA
OTH	AAAGCC
RSTO	AAAGCG
S0	AAAGCT
S1	AAAAGA
S2	AAAAGG
S3	AAACGA
REJ	AAACGC
RSTR	AAACGG
RSTOS0	AAACGT
SH	AAAAAC

Protocol:

Table 3 *Network protocols and their corresponding nucleotide sequence*

Protocol	Nucleotides
TCP	TTTAAT
UDP	TTTGAC
ICMP	TTTGAT

Service:

Table 4 *Network services and their corresponding nucleotide sequence*

Service	Nucleotides
HTTP	CCCTGC
DOMAIN_U	CCCTGT
ECR_I	CCCGAA
SMTP	CCCGAG
FINGER	CCCCAA
FTP_DATA	CCCCAG
POP_3	CCCGGA
AUTH	CCCGGC
ECO_I	CCCGGT
TELNET	CCCCAC
NTP_U	CCCCAT
URP_I	CCCATA
OTHER	CCCATC
PRIVATE	CCCATT
VMNET	CCCCTA
BGP	CCCCTC
Z39_50	CCCCTG
FTP	CCCCTT
SSH	CCCTTA
WHOIS	CCCTTG
DOMAIN	CCCAAG
GOPHER	CCCATG
REMOTE_JOB	CCCTTC
RJE	CCCCCA
CTF	CCCCCG
LINK	CCCCCT
SUPDUP	CCCAGC
ISO_TSAP	CCCGTC
HOSTNAMES	CCCGTG
CSNET_NS	CCCGTT
POP_2	CCCTGA
SUNRPC	CCCTAG
UUCP_PATH	CCCTAA
NNTP	CCCGCA
IMAP4	CCCGCC
SQL_NET	CCCGCG
LDAP	CCCGCT
HTTP_443	CCCAGA
EXEC	CCCAGG
NETBIOS_DGM	CCCCGA
LOGIN	CCCCGC
SHELL	CCCCGG
PRINTER	CCCCGT
EFS	CCCAAC
COURIER	CCCAAT
NETBIOS_SSN	CCCGAC
KSHELL	CCCGAT
DISCARD	CCCAGT
DAYTIME	CCCTCA
SYSTAT	CCCTCC
NETSTAT	CCCTCG
TIME	CCCTCT
NAME	CCCACA
KLOGIN	CCCACC
IRC	CCCACG
X11	CCCACU
ECO	CCCTGG
ICMP	CCCTAC
MTP	CCCTAT
NETBIOS_NS	CCCGTA
NNSP	CCCTGCTGC
PM_DUMP	CCCTGTTGT
TFTP_U	CCCGAAGAA
TIM_I	CCCGAGGAG
UUCP	CCCCAACAA
AOL	CCCCAGCAG
HARVEST	CCCGGAGGA
HTTP_2784	CCCGGCGGC
HTTP_8001	CCCGGTGGT
RED_I	CCCCACCAC
URH_I	CCCCATCAT

3. Dynamic Parameters: There are parameters that dynamically have their values changed according to specific network activity type, these parameter values are mapped based on their values' data types as the following:

   Integers: Integer number is mapped into DNA nucleotides by taking each digit from left to right and representing it with corresponding sequence as the following:

   Table 5 *Integer digits and their corresponding nucleotide sequence*

Digits	Nucleotides
0	AGT
1	TCA
2	TCC
3	TCG
4	TCT
5	ACA
6	ACC
7	ACG
8	ACT
9	TGG

   Reals: The real number is mapped into DNA nucleotides as the integer number in addition to the floating point that mapped to the following:

   Table 6 *Floating point and its corresponding nucleotide sequence*

Symbol	Nucleotides
.	TAC

   Booleans: The logical values which are true/false are mapped to the following nucleotides:

   Table 7 *Logical values and their corresponding nucleotide sequence*

Symbol	Nucleotides
True	TAT
False	GTA

   Fig.1 is sample of normal network activity that is encoded using above encoding methodology and its length is 150 nucleotides.

AGUUUUAAUCCCUGCAAAGCAUCCUCAACAUCUACAAGUACGACC

AGUAGUGUAGUAAGUUAUAGUGUAGUAAGUAGUAGUAGUAGUGU

AGUAUCAUCAAGUAGUAGUAGUUCAAGUAGUAGUAGUAGUAGUA

GUAGUAGUAGUAGUAGU

Fig.1 *Sample of encoded sequence for normal activity*

3.4 Genetic Algorithm Operators

Genetic algorithm has some operators that can be customized and used to optimize the selection for target solution. Using genetic algorithm has many benefits in enhancing the optimization for intrusion detection system behavior [3].

In our case we use genetic algorithm operators to generate and find DNA signature that is considered the best sequence in last generation that has maximum summation for alignment values against randomly selected normal sample activities. The individuals for initial population are sequences for which each one is randomly built using same encoding methodology.

We select a randomly sample from normal sequences; these sequences are aligned with each individual in the population so as to find the fitness value for that individual. After specified

number of generations, the genetic algorithm selects the sequence form the last population that has best fitness value from aligning it with all the normal sample sequences. That sequence is the signature sequence for system normal behavior.
In the following we introduce the genetic algorithm operators we use:

3.4.1 Fitness Function

Each individual sequence in the population is considered as candidate DNA signature and aligned with randomly selected normal sample sequences. The fitness value for each individual is the summation for its alignment scores with all normal sequences from randomly selected sample. The best child is chosen which has maximum fitness value.

3.4.2 Mutation Function

We randomly select two individual sequences from the current population and generate random list of integers that represent the site indices of the mutant positions in the selected individual sequences. For each randomly selected site position we interchange the site position characters between the two sequences. The maximum position limit value in the random list is the length of the smaller sequence.
For example:
Individual Sequence 1:
ATCGCCGTACCCGGTAAATTTT

Individual Sequence 2:
CGCTTACAAGGCCCC

Random List of interchange positions: 2, 6, 10

New mutant child 1:
A**G**CGC**A**GTA**G**CCGGTAAATTTT

New mutant child 2:
C**T**CTT**C**CAA**C**GCCCC

3.4.3 Crossover Function

We randomly select two individual sequences from the current population in addition to generate random number which represents the site position that we will cross over the selected sequences around it. After specifying the position of cross over site, we interchange the sequences parts around it from the two sequences and that generates two new sequences which considered the crossover contribution to the next population. The maximum site position available for crossover operator is half the length of the smaller sequence.
For example:
Individual Sequence 1:
TAGCTGGTAGGCTTTAAAA

Individual Sequence 2:
TGCTCAATGCGCTTGAGTGAAACGGT

Random crossover site position: 10

New interchanged child 1:
TAGCTGGTAG**GCTTGAGTGAAACGGT**

New interchanged child 2:
TGCTCAATGC**GCTTTAAAA**

3.5 Experiment

Testing phase consists of steps to test our suggested intrusion detection system by encoding 5,000 testing activities into DNA sequence then to randomly select the target testing sequences to calculate the accuracy through aligning them against signature sequence. We label each sequence with the result output from detection system as normal or attack label based on the comparison with normal threshold value, if the alignment value is greater than threshold value then it is considered as normal, otherwise, it is an attack.

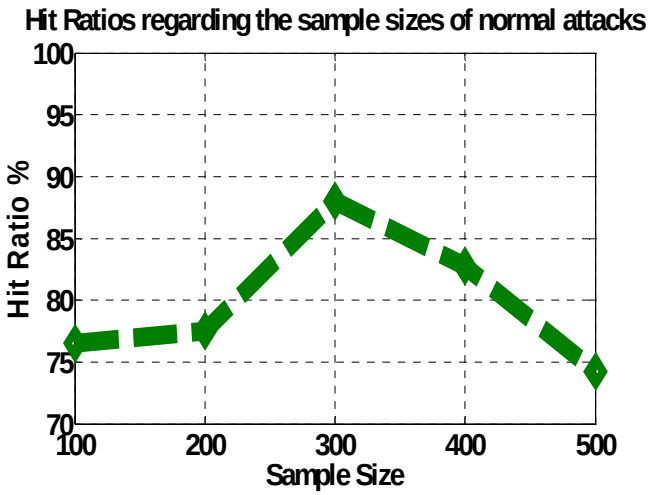

Fig.2 *Variation of Hit ratio (average=79.79) against sample size*

We notice that when increasing the sample size from 100 to 300, the hit ratio values are increased till it starts goes down starting from sample size 300.

Table 8 *Hit ratios for attack categories*

Methods/Attacks	**Probe**	**DOS**	**U2R**	**R2L**	**Normal**
DNA Sequence alignment using Genetic algorithm	57.28%	51.83%	43.10%	24.20%	79.79%

The R2L category has the lowest hit ratio among the other categories as in table 8.

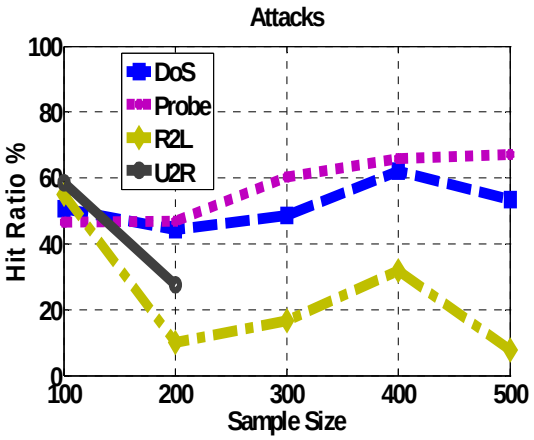

Fig.3 *Hit ratios of attacks and their relationship with sample size*

All categories have low hit ratios in the first 200 sample size then start getting their hit ratios obviously raised after sample size 300.

Table 9 *Normal signature length versus Normal sample size*

Normal Sample	**Normal Signature**
100	1551
200	2013
300	2334
400	2658
500	3297

Table 9 shows the normal signature length comparison against sample size. First column represents the sample size of randomly selected normal sequences from dataset, and second column represents the normal signature sequence length in characters unit. The length of normal signature length is measured by how many nucleotides characters the sequence has. On the contrary the sample size represents how many normal sequences the sample has, and each normal sequence belongs to this sample can have different nucleotides sequence length than others in same sample. Logically, for samples of larger size there is high probability to include many normal activities in the normal signature so the signature sequence length would have its length increased while the normal sample size get increased too.

There is no specific rate for the increase in signature length, since for moving from sample size to the next one which is increased by 100, the increase in signature length differs.

For example, when moving from sample size 100 to 200, the difference in signature length is 462, but when moving from size 400 to 500 sequences, the difference is 639.

Table 10 *Relation between Normal sequence length, signature length and*

number of selected sequences

Sample Size	Sequence length	Signature length	# Marked sequences
100	163	1551	9
200	169	2013	12
300	171	2334	14
400	172	2658	16
500	173	3297	20

In Table 10, the first column is the sample size for normal activities that we randomly select from data set.
The second column represents the average length in characters unit, for the appended normal sequences that we select from the sample to build the normal signature sequence. The third column represents the length in characters unit for the built normal signature sequence in each sample size case. Last column contains the number of marked and selected normal sequences from normal sample list, which they are appended to build the normal signature. To explain the signature building method in more details, let's take the first record of sample size 100 as example. We notice that there are only 9 normal sequences from the whole 100 sample size are selected and appended to represent the signature. We can expect the appended signature length from those 9 sequences to be 1467 characters, in case the average length of the normal sequences from sample size 100 is 163. The actual normal signature length is 1551 nucleotide characters which is different than estimated length 1467 nucleotide characters, because the difference in the selected normal sequences lengths. Quality of normal activities which formulate the sample list affects the number of marked normal sequences and consequently the length of normal signature sequence. When there are many similar network sequences in the sample list that leads to select fewer sequences to formulate the signature and consequently generates shorter signature sequence. On the contrary when there are many different sequences in the sample list, means there is high probability for marking more sequences to formulate the signature and so we have longer signature sequence. Threshold value is important since it actually refers to the normal signature sequence and normal behavior. We do a measurement to check the difference for threshold values while increasing normal sample size, and the result as in Table 11. First column represents sample size of source normal activities list, and second column represents threshold value, which used as the minimum level marker for activity to be considered as normal.

Table 11 *Relation between Sample size and Threshold value*

Sample Size	Threshold value
100	223.6
200	231.8
300	234.5
400	235.8
500	236.6

In Table 11 the threshold value represents the average of the alignment values between the signature and the normal activities sequences of the sample, so in first sample which has 100 normal activities, we calculate the average of the alignments between the signature and each sequence for the 100 normal activities.
Table 12 is a comparison for the average hit ratios for different attack types. We notice that "apache2" attack has the best average hit ratio between the other attack types.

Table 12 *Average hit ratio values for variety of attacks*

Attack	Average hit ratio %
Warez master	95.04
Smurf	80.00
Snmp get	73.84
Mail bomb	52.00
Process table	79.78
IP sweep	37.71
http tunnel	59.31
apache2	**99.00**
Neptune	50.30
Mscan	81.20
Back	64.61
Satan	63.51

Buffer overflow	95.45
Guess passwd	42.00
Rootkit	83.08
Ftp Write	93.33
Multi hop	90.00
Pod	53.33
Tear drop	83.34
Nmap	70.00
Portsweep	96.83
Named	77.65
Xlock	97.78
Sendmail	85.34
Saint	98.76
Xterm	81.54
Ps	72.50

When Smurf attack type has average hit ratio as 80.00 % means that our detection system successfully detected 80 activities from each 100 activities tested by the system and related to Smurf attack type.

Our system generates threshold value as metric for the detection accuracy. Table 13 lists both threshold values and corresponding normal signature length for different normal sample sizes. The normal sample list is a list of normal type network activities that exist in KDD data audit files, those activities are selected randomly with count equal to the specified sample size. In case of sample size 300, we randomly selected 300 normal type activities from KDD data files, and consider these normal activities as a source list to do 5 experiments of training phase execution. We collected 5 sample lists with sizes 100,200,300,400, and 500 sequences. Each sample list activities are randomly selected from KDD normal activities. For sample size 100 of normal activities, we do an experiment on each of 5 randomly selected normal sample lists with size 100 for each. We generated a 5 pairs each of which consist of signature length and threshold value from each experiment. We repeat that experiment for the other sample size lists of normal activities with sizes 200, 300, 400, and 500 sequences. First and second columns are the sample size labels which represent the sample lists size of the source normal activities. Third column "Normal Signature Length/Characters" represents normal signature length of the generated normal signature from each of 5 experiments on the corresponding sample list size. Last column "Threshold value" represents threshold value which is corresponding to each generated normal signature from each experiment in same sample size.

Table 13 *Relationship between threshold values and normal signature length group by different normal sample sizes*

Sample Size		**Normal Signature Length/Characters**	**Threshold value**
	100	315	211.8
		648	205.5
		804	206.9
		819	212.3
		831	209.9
	200	1764	212.1
		2256	211.3
		2463	214.4
		2736	212.4
		4344	213.1
	300	1164	214.3
		1407	195.7
		1809	214.5
		2961	218.1
		3330	206.8
	400	636	206.9
		969	209.5
		2055	210.7
		3228	213.7
		3360	210.2

	500	816	207.7
		1464	205.5
		1884	207.4
		2553	209.8
		4743	205.6

Important note to mention is the noticeable variation among the signature length for same sample size experiments. For example in sample size 500 sequences, first experiment has 816 characters – signature length, on the other hand it has 4743 characters – signature length in fifth experiment. The reason for that significant variation in length difference is the quality and similarity ratio among normal activities in same sample size experiment. The significant of second column, is to highlight the effect of the signature length which decreases the performance of the alignment process during detection method of the observed behavior. The relationship between master signature length and threshold value is important to be studied so as to predict the increase range of the signature length according to sample list size increase.

4 Conclusion

Bioinformatics has sequence alignment methods that help to calculate the similarity degree between system signature sequence and observed network activity sequence. The system decides to label the activity as attack or normal based on comparing threshold value with alignment score. The DNA encoding method has important role to play in detection process since the efficient encoding would decrease the saved sequence size in addition to decrease the time of alignment method execution. The DNA encoding method has an important role to play in detection process since the efficient encoding would decrease the saved sequence size in addition to increase the performance of alignment methods. Many enhancements on this technique can be done by improving the encoding methodology for the network activities into DNA sequence. Intrusion detection has many techniques to be implemented, we explain a technique that encodes network activities into DNA sequences and calculates alignment threshold value then to measure the alignment value for suspected network activity. Our technique uses KDD'99 dataset so as trying to reach a high percentage of accuracy for figuring out the network activity as attack or normal one. The intrusion detection technique using bioinformatics methods that are mentioned in this study has wide area to be enhanced and much more deep work to be done on it by finding more efficient encoding DNA sequence, alignment threshold value calculation algorithm, and DNA signature & testing sequences generation process. To enhance the detection process a new efficient encoding DNA sequence is needed which affords any encoding sequence combination for new values. More enhancements can be done on the normal threshold value calculation algorithm and DNA signature so as to reach a sensitive threshold value that accurately reflects the normal behavior of the system. In this paper we use global alignment method to align the sequences and find similarity score between them based on bioinformatics methods, but with more tuning for alignment methods we can achieve faster algorithms which give accurate and better performance even for long sequences. Using genetic algorithm to enhance the optimization for threshold values and signature sequence may promise with good result for such methods.

References:

[1] KDD Cup 1999: The Fifth International Conference on Knowledge Discovery and Data Mining. http://www.acm.org/sigs/sigkdd/kddcup/index.php?section=1999&method=info

[2] Lane, T., and Brodley, C. E. 1997. Sequence Matching and Learning in Anomaly Detection for Computer Security. In Proceedings of AAAI-97 Workshop on AI Approaches to Fraud Detection and Risk Management, Providence, RI, June 1997, pages 43-49.

[3] Li, W. 2004. Using Genetic Algorithm for Network Intrusion Detection. Proceedings of the United States Department of Energy Cyber Security Group 2004 Training

Conference, Kansas City, Kansas, May 2004, pages 24-27.

[4] Needleman, S. B., and Wunch, C. D. 1970. A general method applicable to the search for similarities in the amno acid sequences of two proteins. Journal of Molecular Biology 48, July 1970, pages 443-453.

[5] Porras, P.A., Ilgun, K., and Kemmerer, R.A. 1995. State transition analysis: A rule-based intrusion detection approach, IEEE Transactions on Software Engineering, vol.21, no.3, pp. 181-199, 1995.

[6] Schonlau, M. and Theus, M., "Detecting Masquerades in Intrusion Detection based on Unpopular Commands," Information Processing Letters, 76, November 2000, pp. 33-38.

[7] Sekar, R. 2002. Specification based Anomaly Detection: A New Approach for Detecting Network Intrusions. Proceedings of the 9th ACM conference on Computer and communications security, Washington, DC, December 2002, pp. 265 – 274.

[8] Smith, T. F. and Waterman, M. S. Identification of common molecular subsequences. Journal of Molecular Biology, 147, June 1981, pages 195-197.

[9] Stallings, W. 2003, Cryptography And Network Security, third edition, ISBN # 0-1309-1429-0, Prentice Hall, May 2003, pp. 566 - 577.

[10] Tian, D., Liu, Y., and Li, B. 2007. A Distributed Hebb Neural Network for Network Anomaly Detection. Springer-Verlag Berlin Heidelberg 2007, I. Stojmenovic et al. (Eds.): ISPA 2007, LNCS 4742, 2007, pp. 314–325.

[11] Traore, I., and Lu W. 2005. A New Unsupervised Anomaly Detection Framework for Detecting Network Attacks in Real-Time. Cryptology and Network Security, ISBN # 978-3-540-30849-2, Springer Berlin / Heidelberg, November 2005, pp. 96-109

[12] Watson, J. D., and Crick, F. 1953. Genetical Implications of the structure of deoxyribonucleic acid. Nature 171, 4361, May 1953, pp. 964-967.

[13] Yeung, D., and Chow, C. 2002. Parzen-Window Network Intrusion Detectors. Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02), 2002, 1051- 4651/02.

[14] Yu, B., Byers, E. J., and Howey, C. 2001. Monitoring Controller's "DNA Sequence" For System Security. ISA Emerging Technologies Conference, Instrumentation Systems and Automation Society, Houston, September 2001.