INTRUSION DETECTION IN COMPUTER NETWORK USING GENETIC ALGORITHM APPROACH: A SURVEY

S. N. Pawar
Associate Professor (E &TC),
Jawaharlal Nehru Engineering College, Aurangabad, MS, India.

ABSTRACT

The intrusion detection problem is becoming a challenging task due to the proliferation of heterogeneous computer networks since the increased connectivity of computer systems gives greater access to outsiders and makes it easier for intruders to avoid identification [1]. Intrusion detection systems are used to detect unauthorized access to a computer system. A number of soft computing based approaches are being used for detecting network intrusions. This paper presents a survey on intrusion detection techniques that use genetic algorithm approach.

KEYWORDS: *Intrusion Detection, Genetic Algorithm.*

I. INTRODUCTION

When a computer system is connected to a network, it goes on a high risk. There are various threats to a computer system such as viruses, intrusions etc. Viruses can be greatly controlled by installing antivirus software and updating its virus files regularly.

Any unauthorized access to the resources of the computer is called intrusion to a computer. Intrusions can be detected by installing intrusion detection system (IDS). Various soft computing techniques such as Genetic Algorithm, Artificial Neural Network, Support vector machine and Fuzzy Logic are used to make an intrusion detection system (IDS). Genetic algorithm (GA) alone or in combination with some other artificial intelligence technique is found to be the most efficient approach for intrusion detection [2-12]. The functions of IDSs are to detect the intrusions, generate the pop up message to the user and take the necessary corrective action.

There are number of limitations to the prevention-based approach for computer and network security [13]. It is probably impossible to build a completely secure system. The prevention-based security philosophy constrains the user's activity and productivity. Hence, intrusion detection systems are designed based on various detection techniques. IDSs are of two types: Anomaly intrusion detection and Misuse intrusion detection [14].

In anomaly IDS, the users' behaviour is compared with a known standard behaviour and any significant deviation from normal behaviour is detected. This approach can be more effective in protection against unknown or novel attacks since no prior knowledge about specific intrusions is required. However, it may cause more false positives because abnormality can be due to a new normal behaviour [15].

The misuse intrusion detection is the most widely used IDS. It uses patterns of known attacks or weak spot of the system to identify known intrusions. The signatures and patterns used to identify attacks consist of various options in the packet like source address, destination address, source and destination ports and even the keywords in the content area of a packet.

The IDSs can also be classified in to two categories based on where they look for intrusions: host-based IDS and network-based IDS. Host-based IDS monitors activities associated with a particular host whereas network-based IDS monitors activities associated with a network [15].

Thus far, I have discussed the motivation of the presented work and the brief overview of the IDS. The rest of the paper is organized as follows. Section II gives an overview of the genetic algorithm

employed in this work. In Section III survey of the work relevant is made. Section IV presents the data sets used by different researchers. Section V presents the results obtained by these researchers and section VI concludes the paper.

II. GENETIC ALGORITHM

GA is the technique which works on the mechanics of natural selection. They are based on the Darwin's theory of survival of the fittest. The main reason behind the design of GA was to abstract and rigorously explain the adaptive processes of natural selection and to design artificial system that retrains two important mechanics of natural systems [16]. The major application of GA is in the area of optimization.

The GA process begins with a set of potential solutions or chromosomes (usually in the form of bit string) which are randomly generated or selected. The entire set of these chromosomes comprises a population. The chromosomes evolve during several iterations or generations. New offspring are generated using the crossover and mutation technique.

Crossover involves splitting two chromosomes and then combining first part of a chromosome with the second part of the other chromosome. Mutation involves flipping one or more bits of a chromosome. The chromosomes are then evaluated using a certain fitness criteria. After the termination criterion is satisfied, the chromosome having the highest fitness is taken as the best solution of the problem [16].

GAs are different from the other search and optimization procedures. They work with a coding of the parameter set, search from a population of points, use payoffs i.e. objective function information and use probabilistic transition rules.

Following are the GA operators which are applied on a population of chromosomes [16].

a. Selection

The selection operator determines which chromosome(s) from the population will be chosen for recombination based on the fitness of the chromosomes. The selected chromosomes are called parents. Such selection methods are: fitness proportion selection, roulette-wheel selection, stochastic universal sampling, local selection and rank selection.

b. Crossover

The parent chromosomes are recombined by one of the crossover methods. It produces one or more new chromosomes(s) called offspring(s). Such methods are: single point crossover, multipoint crossover, uniform and arithmetic crossover.

c. Mutation

New genetic material could be introduced in to the new population through mutation process. This will increase the diversity in the population.

III. USING GA APPROACH IN INTRUSION DETECTION

Different researchers have implemented GA in a different way for network intrusion detection.

Melani J Middlemiss et al. (2003) [17] have used GA for weighted feature extraction with specific application to intrusion detection data. They have implemented a simple genetic algorithm which evolves weights for the features of data set. A *k*-nearest neighbour classifier was used for the fitness function of GA as well as to evaluate the performance of the new weighted feature set.

Wei Li (2004) [18] presents a technique of applying GA to IDSs. After giving a brief introduction to IDS, GA and related detection techniques, he has discussed various implementation details. He has used GA to generate the classification rules which were used to classify normal network connections from anomalous connections. These rules are in if {condition} then {act} form. He encoded chromosomes in integer form but IP addresses are encoded in hexadecimal form. Chromosome population is randomly selected. Population is evolved using crossover and mutation. Effective fitness function is used to check the fitness of each rule. Fittest rules are then used for intrusion detection.

Ren Hui Gong et al. (2005) [15] have used a simple genetic algorithm to derive a set of classification rules from network audit data and the support confidence framework is utilized as fitness function to

judge the quality of each rule. The generated rules are then used to detect or classify network intrusions in a real time environment.

Jiu-Ling Zhao et al. (2005) [19] have presented a novel approach of using clustering genetic algorithms to solve the computer network intrusion detection problem. They described a prototype intelligent intrusion detection system to demonstrate the effectiveness. This system combines two stages in to the process including clustering stage and genetic optimization stage. The algorithm can not only cluster the cases automatically, but also detect the unknown intruded action.

Tao Xia et al. (2005) [20] present a hybrid method based on information theory and genetic algorithm to detect network attacks. Information theory is used to filter the traffic data and thus reduce the complexity. A linear structure rule is used to classify the network behaviour in the normal and abnormal behaviours.

Chi Hoon Lee et al. (2006) [21] presents the novel feature selection method that maximizes class separation between normal and attack patterns of computer network connections. They have focused on selecting a robust feature subset based on the genetic optimization procedure in order to improve a true positive intrusion detection rate.

Saqib Ashfaq et al. (2006) [22] have used a genetic algorithm for generating efficient rules for cost sensitive misuse detection in intrusion detection systems. They have used five most weighted features identified by M.J.Middlemiss et al. [17]. They have designed a GA to identify these features. The algorithm generates if–then rules that identify an attack as well as its category so that appropriate action can be taken in response. This approach is cost sensitive that considers the cost of false alarms for each category of attack separately.

Nalini N. and Raghavendra Rao G. (2006) [23] present a novel method of intrusion detection based on genetic algorithms and principal component analysis. This technique can also be used to detect the class of intrusion. In this paper, they experiment with PCA to reduce the number of features of a TCP connection. This helps in reducing the number of bits required to represent a connection without loss of significant information. They show how network connection information can be modelled as chromosomes and how the parameters in genetic algorithm can be defined in this respect.

Hua Zhou et al. (2007) [24] have used SVM and Genetic Algorithm to increase the classification accuracy. They used GA for feature selection and optimization and then used SVM model to detect intrusions.

Yong Wang et al. (2009) [25] propose a fitness function, an efficient rule generator for denial of service attack. He used GA toolbox provided by MATLAB (R14) for his implementation. He designed the genetic algorithm using 4 m-files. The rules generated are in if {condition} then {outcome} form. The rules generated are suitable for continuously changing misuse detection.

Chen Zhongmin et al. (2009) [26] designed a training algorithm model based on abnormality detection. The proposed experimental model is based on a hypothesis that if variable x appears more times than the desired value, there is possibility of occurrence of abnormality.

Chris Sinclair et al. (2010) [27] have proposed an approach to create rules for an intrusion detection expert system. They employ genetic algorithms and decision trees to automatically generate rules for classifying network connections. They have used genetic algorithms to evolve simple classification rules. The rules are in if {pattern matched} then {generate alert} form. They have used only five attributes such as source IP address, destination IP address, source IP port, destination IP port, and network protocol. They have used the ID3 algorithm to create decision trees. The final rule generated by decision tree is complete with respect to the training data.

S.N.Pawar and R.S.Bichkar (2012) [28] have used genetic algorithm with enumeration technique. The enumeration technique is used while generating random population. It is used to determine the value of each gene.This enumeration technique substantially reduces the computational time required for population generation and yields more appropriate rules.

Table 1 summarizes various GA based approaches used by these researchers and various GA parameters that determine the performance of the genetic algorithm such as chromosome representation, selection, crossover, mutation, population size and generations.

Table 1. The GA approaches and GA parameters used by different researchers for intrusion detection.

Sr. No	Name of the Researcher and year	Approach used	GA parameters					% Detection
			Selection	Cross-over	Mutation	Population	Generations	
1	Melani J Middlemiss et al. (2003)	Simple GA with a KNN classifier	Linear ranking	0.6	0.0075	100	100	NA*
2	Ren Hui Gong et al. (2005)	GA	Fitness Proportionate	0.5	0.02	500	5000	79.8
3	Jiu-Ling Zhao et al. (2005)	Clustering GA	NA	0.75	0.001	120	200	95
4	Tao Xia et al. (2005)	Hybrid method based on information theory and GA	NA	NA	0.01	1000	NA	99.2
5	Chi Hoon Lee et al. (2006)	GA	NA	0.6	0.05	30	50	62.9
6	Saqib Ashfaq et al. (2006)	GA	Fitness Proportionate	NA	NA	200	50	96.4
7	Nalini N. and Raghavendra Rao G (2006)	GA and PCA	NA	NA	NA	NA	NA	93.1
8	Hua Zhou, Xiangru Meng, Li Zhang (2007)	GA and SVM	Fitness Proportionate	0.8	0.05	100	300	98.9
9	Yong Wang et al. (2009)	GA	NA	NA	NA	NA	100	95.4
10	Chen Zhongmin et al. (2009)	GA	Fitness Proportionate	0.500	0.001	100	NA	NA
11	Chris Sinclair et al. (2010)	GA and decision trees	NA	NA	NA	NA	NA	NA
12	S.N.Pawar and R.S.Bichkar (2012)	GA with enumeration	Fitness Proportionate	0.5	0.01	300	2000	98

(*Not Available)

IV. DATA SET USED

4.1 DARPA 1998

All the researchers have implemented their genetic algorithm on the offline data such as DARPA1998 data or KDD CUP 99 data. MIT Lincoln Laboratory, under Defence Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory (AFRL) sponsorship, has collected and distributed the first standard data for evaluation of computer network intrusion detection systems.

This Data is DARPA 1998 data [29]. It consist of tcpdump and BSM list files. Each line in a list file corresponds to a separate session. Each session corresponds to an individual TCP/IP connection between two computers. The first nine columns in list files provide information which identifies the TCP/IP connection.

4.2 KDD CUP 99

KDD CUP 99 data [30] comes from the KDD Cup competition. This data is part of the data collected from the MIT Lincoln Labs 1998 DARPA Intrusion Detection Evaluation Program and is considered benchmark data for evaluating intrusion detection systems. The data is available in full (743M of network connections), or in a number of smaller data sets.
Table 2 below summarizes the various datasets used by the researchers

Table 2. Data set used by different researchers.

Sr. No	Name of the Researcher and year	Data set used
1	Melani J Middlemiss et al. (2003)	KDD CUP 99
2	Ren Hui Gong et al. (2005)	DARPA 1998
3	Jiu-Ling Zhao et al. (2005)	KDD CUP 99
4	Tao Xia et al. (2005)	KDD CUP 99
5	Chi Hoon Lee et al. (2006)	KDD CUP 99
6	Saqib Ashfaq et al. (2006)	KDD CUP 99
7	Nalini N. and Raghavendra Rao G (2006)	KDD CUP 99
8	Hua Zhou, Xiangru Meng, Li Zhang (2007)	KDD CUP 99
9	Yong Wang et al. (2009)	KDD CUP 99
10	Chen Zhongmin et al. (2009)	DARPA 1998
11	Chris Sinclair et al. (2010)	DARPA 1998
12	S.N.Pawar and R.S.Bichkar (2012)	DARPA 1998

V. RESULTS

Table 1 gives the % detection obtained by using the various GA based approaches. It is evident that Tao Xia et al. (2005) [20] who used hybrid method based on information theory and genetic algorithm got extremely good detection rates (99.25%).
S.N.Pawar and R.S.Bichkar (2012) [28] who used enumeration technique in a GA based rule generation have got good detection rates (98.06%).
Hua Zhou et al. [24] used GA for feature selection and SVM model for classification have also got good results (98.97%).
In all, the intrusion detection results obtained using GA approach are better or are comparable to the results obtained by using other soft computing techniques.

VI. CONCLUSION

Genetic algorithm is found to be one of the efficient technique in network intrusions detection. It is successfully implemented in IDS either to generate the classification rules or to select the appropriate features.
Middlemiss [17], Lee [21] and Hua Zhou et al. [24] have used GA for the selection of appropriate features whereas Gong et al. [15], Saqib Ashfaq et al. [22], Tao Xia et al. [20], Yong Wang et al. [25], Sinclair et al. [27] and S. N. Pawar et al.[28] have used GA for the generation of classification rules.
A hybrid system using GA can be a better solution to decrease the false positive rate [17, 19, 20, 23 and 24].
Using enumeration in a GA based IDS reduces the search space and yields more accurate results [28].

ACKNOWLEDGMENTS

The author wishes to extend his sincere thanks to Dr. R. S. Bichkar, Professor (E&TC), G. H. Raisoni College of Engineering & Management, Pune, MS, India, for the review and valuable suggestions in writing this paper.

REFERENCES

[1]. Biswanath Mukherjee, L. Todd Herberlein and Karl N. Levitt, "Network Intrusion Detection", IEEE Network, 8 (3): 26-41,May/June 1994.
[2]. Srinivas Mukkamala and Andrew H. Sung, "A Comparative Study of Techniques for Intrusion Detection*", Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence* (ICTAI'03), IEEE, 2003.
[3]. Jing Xiaopei, Wang Houxiang, Han Ruofei and Li Juan, "Improved Genetic Algorithm in Intrusion Detection Model Based on Artificial Immune Theory," IEEE, 2009.
[4]. Hua Zhou, Xiangru Meng and Li Zhang, "Application of Support Vector Machine and Genetic Algorithm to Network Intrusion Detection," IEEE, 2007.
[5]. Shingo Mabu, Ci Chen, Nannan Lu, Kaoru Shimada and Kotaro Hirasawa, "An Intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming," *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*. IEEE, 2010.
[6]. Yu-Ping Zhou, Jian-An Fang and Dong-Mei Yu, "Research on Fuzzy Genetics-Based Rule Classifier in Intrusion Detection System," *International Conference on Intelli- gent Computation Technology and Automation*. IEEE, 2008.
[7]. A.T. Haghighat, M. Esmaeili, A. Saremi and V. R. Mousavi**, "**Intrusion Detection via Fuzzy-Genetic Algorithm Combination with Evolutionary Algorithms", *International Conference on Computer and Information Science* (ICIS 2007), IEEE, 2007.
[8]. Siva S. Sivatha Sindhu, S. Geetha, Siva. S. Sivanath and Dr. A. Kannan, "A Neuro-genetic ensemble Short Term Forecasting Framework for Anomaly Intrusion Prediction", IEEE, 2006.
[9]. Dong Seong Kim, Ha-Nam Nguyen and Jong Sou Park, "Genetic Algorithm to Improve SVM-based Network Intrusion Detection System", *19th International Conference on Advanced Information Networking and Applications* (AINA'05), IEEE, 2005.
[10]. Khaja Mohammad Shazzad and Jong Sou Park, "Optimization of Intrusion Detection through Fast Hybrid Feature Selection", *Sixth International Conference on Parallel and Distributed Computing, Applications and Technologies* (PDCAT'05), IEEE, 2005.
[11]. Yongxuan Zhu, Xin Shan and Jun Guo, "Modified Genetic Algorithm-based Feature Subset Selection in Intrusion Detection System", *Proceedings of ISCIT*, IEEE, 2005.
[12]. Yu-Ping Zhou, Jian-An Fang and Dong-Mei Yu, "Research on Fuzzy Genetics-Based Rule Classifier in Intrusion Detection System", *International Conference on Intelligent Computation Technology and Automation*. IEEE, 2008.
[13]. A.Vesely and D.Brechlerova, "Neural Networks in Intrusion Detection Systems", AGRIC. ECON. CZECH, 50, 2004 (1): 35-39.
[14]. S.Owais, V.Snasel, P.Kromer and A. Abraham, "Survey Using Genetic Algorithm Approach in Intrusion Detection Systems Techniques", *7th Computer Information Systems and Industrial Management Applications*, 2008, IEEE , pp.300-307.
[15]. Ren Hui Gong, Mohammad Zulkernine and Purang Abolmaesumi, "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection", SNPD/SAWN'05, IEEE, 2005.
[16]. David E. Goldberg: Genetic Algorithms in Search, Optimization and Machine Learning. Pearson Education, seventh reprint, 2004.
[17]. Melanie Middlemiss and Grant Dick, "Weighted Feature Extraction Using a Genetic Algorithm for Intrusion Detection", *2003 Congress on Evolutionary Computation* (cec-03) 2003, pp.1669-1675.
[18]. W. Li, "Using Genetic Algorithm for Network Intrusion Detection", *Proceedings of the United States Department of Energy Cyber Security Group*, 2004.
[19]. Jiu-Ling Zhao, Jiu–Fen Zhao and Jian –Jun Li, "Intrusion Detection Based on Clustering Genetic Algorithm", *International Conference Based on Machine Learning and Cybernetics*, IEEE, Guangzhou, 2005, pp.3911-3914.
[20]. Tao Xia, Guangzhi Qu, Salim Hariri and Mazin Yousif, "An efficient Network Intrusion Detection Method Based on Information Theory and Genetic Algorithm", IEEE, 2005.
[21]. Chi Hoon Lee, Sung Woo Shin and Jin Wook Chung, "Network Intrusion Detection Through Genetic Feature Selection", SNPD, IEEE, 2006.

[22]. Saqib Ashfaq, M.Umar Farooq and Asim Karim, "Efficient Rule Generation for Cost- Sensitive Misuse Detection Using Genetic Algorithms", IEEE, 2006.
[23]. Nalini N and Raghavendra Rao G., "Network Intrusion Detection via a Hybrid of Genetic Algorithms and Principal Component Analysis", IEEE, 2006.
[24]. Hua Zhou, Xingu Meng and Li Zhang, "Application of Support Vector Machine and Genetic Algorithm to Network Intrusion Detection", IEEE, 2007.
[25]. Yong Wang, Dawu Gu, Xiuxia Tian and Jing Li, "Genetic Algorithm Rule Definition for Denial of Services Network Intrusion Detection", *International Conference on Computational Intelligence and Natural Computing*, IEEE, 2009, pp.99-102.
[26]. Chen Zhongmin, Feng Jianyuan, Xu Sheng and Xu Renzuo, "The research of Intrusion Detection Technology Based on Genetic Algorithms", *International Conference on Net-works Security, Wireless Commu-nications and Trusted Computing*, IEEE, 2009.
[27]. Sinclair Chris, L. Pierce, and S. Matzner, "An Application of Machine Learning to Network Intrusion Detection", *Annual Computer Security Applications Conference*, 1999, pp. 371-377.
[28]. S. N. Pawar and R. S. Bichkar, "Using Enumeration in a GA based Intrusion Detection", *International Journal of Computer Applications* (IJCA), October, 2012.
[29]. MIT Lincoln Laboratory, DARPA datasets, MIT, USA. (http://www.ll.mit.edu/mission/communications /ist /corpora /ideval/data/1998data.html)
[30]. KDD CUP 99 data set (http://kdd.ics.uci.edu/ databases/kddcup99/kdd cup)

AUTHOR

S. N. Pawar was born in Parbhani, India, in1969. He received the Bachelor in Electronics Engineering degree from Marathwada University, Aurangabad, in Year 1993 and the M.Tech in Electronics Product Design degree from Dr. B. A. M. University, in Year 2000. He is currently pursuing the Ph.D. degree with the Department of Electronics and Telecommunication Engineering, S.G.G.S. College of Engineering and Technology, Nanded. His research interest include Genetic algorithm, Network security.