Indonesian Journal of Electrical Engineering and Computer Science
Vol. 17, No. 2, February 2020, pp. 1060~1065
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v17.i2.pp1060-1065

Network intrusion detection system using immune-genetic algorithm (IGA)

Hamizan Suhaimi[1], Saiful Izwan Suliman[2], Ismail Musirin[3], Afdallyna Harun[4], Roslina Mohamad[5], Murizah Kassim[6], Shahrani Shahbudin[7]
[1,2,3,5,6,7]Faculty of Electrical Engineering, Universiti Teknologi MARA, Malaysia
[4]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia

Article Info

Article history:

Received May 12, 2019
Revised Aug 14, 2019
Accepted Aug 28, 2019

Keywords:

Computer network systems
Genetic algorithm
Immune-genetic algorithm
Intrusion detection system

ABSTRACT

Network security is an important aspect in maintaining computer network systems and personal information from being illegally accessed by third parties. The major problem that frequently occurs in computer network systems is the failure in detecting possible network-attacks. Apart from that, the process of recognizing the type of attack that occurs is very crucial as it will determine the elimination process that should take place to counter the intrusion. This paper proposes the application of standard Genetic Algorithm (GA) that combines with immune algorithm process to enhance the computer system's capability in recognizing possible intrusion occurrence in a computer system. Simulation was conducted numerous times to test the effectiveness of the proposed intrusion detection system by manipulating the parameter values for genetic operators utilized in GA. The effectiveness of the proposed method is shown in the gathered results and the analysis conducted further supports and proves that Immune Genetic Algorithm (IGA) has the capability to predict the occurrence of intrusion in computer network.

Corresponding Author:

Saiful Izwan Suliman,
Faculty of Electrical Engineering,
Universiti Teknologi MARA,
40450 Shah Alam, Selangor, Malaysia.
Email: saifulizwan@uitm.edu.my

1. INTRODUCTION

In line with the development of communication technology, the data security in a computer system is now more vulnerable to threats that could be possibly misused for a wrong reason. In order to deal with this situation, a good and efficient mechanism must be developed and employed as a preemptive method to protect the valuable information available in a computer system [1-4]. The unauthorized access of sensitive information does not only occur externally from a network (Internet), but also internally (intranet). This study primarily focuses on developing a system that can detect the occurrence of unauthorized activity in a computer network and from the intrusion connections detected the type of the attack will be identified. These detection and identification processes were mainly modelled based on metaheuristic approach which is Genetic Algorithm (GA) [5-6]. This algorithm alone is a powerful method in optimization area. However, hybridization of its algorithm with other method will explore the possibility of enhancing the performance in many complex problems. In this study, proliferation processes through cloning together with affinity maturation process which are extracted from Artificial Immune System (AIS) were incorporated into the standard GA for pattern recognition purpose. In the computer system nowadays, most software and

applications are executed independently with no around-the-clock dedicated person in charge to monitor the system. Furthermore, online systems are connected to the Internet all the time. This condition in addition to the restricted resources in a network highlights the need to have an efficient system that could monitor and detect the possibility of intrusions and unauthorized access [7-10].

Network Intrusion Detection System (NIDS) is a pattern matching which is called misuse-based IDS or signature-based IDS. It is able to determine and recognize the known patterns that have been stored in IDS database [11-13]. Based on National Institute of Standards and Technology (NIST) organization, IDS classified into three different categories: Host-based IDS, Network-based IDS and Vulnerability-assessment IDS [14]. An intrusion detection technique proposed in this paper was developed based on two metaheuristic approaches which are GA and AIS. This proposed method termed as Immune-Genetic Algorithm will measure the differences of anomalies and normal connection using a proposed fitness function based on the concept of Euclidean Distance [15]. An Artificial Immune System based on GA was proposed by A. A. Amira, A. E. Hassanien and A. S. Mostafa in January 2012. In their study, they focus on finding any possible approaches that could be utilized to increase the probability of detecting attacks. Their proposed techniques use an algorithm known as deterministic-crowding Niching technique. It was implemented in order to generate detectors for network anomaly intrusion detection system [16]. The benefits of these techniques are that there are no additional parameters needed plus it is simple and fast [17]. GA with different approach fitness function known as Reward-Penalty was proposed by A. Firas and N. Reyadh in 2012 [18]. This technique evaluates the population of chromosomes and penalty will be applied to bad chromosomes and give reward to the good one. The outcomes prove that proposed fitness function works properly and able to produce good results. Artificial Immune System (AIS) as a detection for network intrusion was investigated by S.I. Suliman *et al*. [19]. By using classification method, combination of different connection features can be grouped thus anomalies and legal connections can be distinguished easily. The proposed method has successfully produced good detection with high success rate.

2. RESEARCH METHOD

A network intrusion occurs when an unauthorized user sneak into a computer system for any reason [20]. This could lead to security breach such as sensitive information leaking as well illegal use of resources [21-22]. The proposed method introduces the use of hybrid Genetic Algorithms and Artificial Immune System (AIS) to detect any intrusion that occurs in a computer network. Genetic Algorithm is one of Evolutionary Algorithm (EA) approaches in the field of Computational Intelligence. Optimization, classification, prediction, recognition and automatic model design are examples of EA applications that are widely used to solve complex engineering problems [23]. In recent years, Differential Evolution (DE) has become most prominent among other EAs in solving optimization problems [24-25]. Besides GA, DE also had been used for scalarizing functions on multi-objective problem to analyze its effect [26]. In IDS, DE widely use for AIS as an optimization technique where it was inspired by biological immune system [27]. It belongs to the computational intelligence family. The proposed method in this paper is investigated and utilized to detect anomalies based on identified features in computer network environment. The fundamental steps involved in this algorithm are the process of natural selection, biological evolution and also genetic recombination. Three main steps involved in the algorithms shown in Figure 1.

Step 1: Separate main dataset into training and testing based on predetermined probability
Step2: Training process
Step 3: Testing process

Figure 1. Three main steps for the proposed Network Intrusion Detection System (NIDS)

2.1. Separating Dataset

The first step involves the random selection process of data to be divided into 2 data sets, namely training data set and testing data set. Each row of data is known as a chromosome, which contains features of an attack type. For the selection, probability of selection is set as benchmark to see the effectiveness of data selection towards the number of attacks that can be detected by the simulation IDS. Three different probabilities were set which is 0.2, 0.3, and 0.5 respectively for selection. By using pre-set probability value, 90% from main data sets will be selected as training data.

2.2. Training Process

In the training process, a number of steps were executed iteratively in order to produce the final population of good quality chromosomes. These chromosomes will be used during the testing process for the purpose of detecting anomaly activity. The steps undertaken during the process are:

2.2.1 Generation of an Initial Population

One hundred chromosomes representing candidate solutions were generated randomly based on the amount of time between time point and the clock's epoch. Each of the generated chromosomes will have forty-one features which refer to the properties of a network connection.

2.2.2 Fitness Function Evaluation

Fitness function is used to evaluate the quality if the generated candidate solutions. In this study, the proposed fitness function is given by the following formula [1]:

$$F = \sqrt{\sum_{i=1}^{n}(P_i - Q_i)^2} \quad (1)$$

From (1), Pi indicates the data of single chromosome while Q_i is the value of an individual training data. The difference between these two values will be squared to get a positive value. The summation of differences for all features will be squared root and measured as *F*, Fitness value. The proposed fitness function was based on Euclidean Distance concept where the smallest distance between two points is considered as the best value.

2.2.3 Clone, Crossover, and Mutation

The probability of getting good-quality solutions can be increased by multiplying the best chromosome, and identified the iteration. This step is known as cloning process. Then the properties in the cloned chromosomes will be altered by performing crossover process between two selected chromosomes. This crossover step emulates the reproduction process in which a child is produced by combining genes from both parents. The genes of the offspring produced will be slightly modified by using mutation process. However, this step will be performed based on predetermined probability value. These recombination of genetic and mutation processes represent genetic evolution to produces new chromosomes which could evolve to be better chromosomes. This process is similar with biological development concepts for human beings and living things. With new chromosomes development, it was expected that it could increase the fitness value of new mutated chromosomes. New population will be generated back by combining top 30 previous random chromosomes, top 30 fitness of mutated chromosomes and 40 new chromosomes by generated randomly. New combination of 100 chromosomes will undergo the same processes from fitness calculation to mutation process iteratively. The final population of chromosomes produced will go through the next phase, which is testing process.

2.3. Testing Process

Figure 2 shows the steps taken during the testing process. This phase will determine whether the testing data is a normal connection or attack connection. As the proposed algorithm is a supervised learning method, the step will also identify the type of attack for each testing data based on outcome of the training process. This step will utilize the same fitness function as discussed in Section 2.2.2. As shown in (1) will be utilized to select the chromosome for each testing data.

Step 1: Each of the testing data is presented to the final population of chromosomes.
Step 2: Fitness value is calculated to determine the type of network connection
Step 3: Calculate the true positive rate of the prediction for all testing data

Figure 2. Steps in testing process

The type of attack for the best chromosome will then be selected as the type of attack for the testing data. If it is similar as the known attack for that particular testing data, then it is considered as a successful prediction. This process is repeated for all testing data. Statistical analysis was also conducted to analyze the overall performance of the prediction produced in order to investigate its efficiency in this task.

3. RESULTS AND ANALYSIS

One of the aspects that could determine the efficiency of the proposed method in this network intrusion detection system is the data separation step for training and testing dataset. This study focuses on three different probability values used whether to include a particular connection data in the testing or training dataset. Each probability value has the potential to influence the Success Detection Rate (SDR) for three main attacks investigated in this study. The simulation program for each probability value was run twenty times and the results obtained were recorded. Statistical analysis was then conducted to see its average, standard deviation and variance values. The results are shown in Table 1, 2, and 3 for 0.2, 0.3, and 0.5 probability of random selection respectively.

Table 1. Results for Experiments with 0.2 Probability of Random Data Selection

Type of Attack	Number of Data	The Best Result	The Worst Result	Average	Standard Deviation, Σ	Variance, σ^2
SnmpGetAttack	774	774 (100%)	517 (66.796%)	731.35 (94.49%)	71.178	5066.308
GuessPassword	437	372 (85.126%)	16 (3.661%)	172.8 (39.542 %)	116.976	13683.385
MailBomb	500	500 (100%)	455 (91%)	482.95 (96.59%)	16.672	277.956

Based on the results gathered in Table 1 which employs selection rate of 0.2, IGA has managed to record 100% accurate prediction for both SnmpGetAttack (774 testing data) and MailBomb (500 testing data) at least once in 20 runs. Meanwhile, the highest accurate prediction for GuessPassword is 85% which represents 372 out of total 437 testing data. The highest average value in twenty runs is the detection of Mailbomb attack which is 96.59% followed by SnmpGetAttack and Guesspassword with 94.49% and 39.542% respectively. The differences between all twenty runs were analyzed using standard deviation, which will determine the total differences between the mean value and the obtained prediction. The lowest standard deviation was observed for MailBomb with 16.672, followed by SnmpGetAttack and GuessPassword. This means that results produced from 20 runs for MailBomb attack is more consistent as compared to the other two attacks.

Table 2. Results for Experiments with 0.3 probability of Random Data Selection

Type of Attack	Number of Data	The Best Result	The Worst Result	Average	Standard Deviation	Variance
SnmpGetAttack	774	774 (100%)	709 (91.602%)	682.6 (88.191 %)	17.295	299.117
GuessPassword	437	272 (62.24 %)	32 (7.323 %)	177.95 (40.721 %)	66.724	4452.092
MailBomb	500	500 (100%)	429 (85.8 %)	477.3 (95.46 %)	27.595	761.484

For program with 0.3 as its probability value, the results obtained are as shown in Table 2. All testing data for SnmpGetAttack and MailBomb are correctly predicted by IGA with 100% accuracy. The proposed method managed to accurately predict 272 testing data for GuessPassword which represents 62.24%. Between SnmpGetAttack and MailBomb, IGA is more consistent in producing almost similar results in 20 runs for the former with standard deviation of 17.295 as compared 27.595 for MailBomb.

Table 3 below record the results for program which utilizes 0.5 probability value for the selection process. Table 3 illustrates that for SnmpGetAttack and MailBomb, our proposed method has again perfectly predicted all testing data allocated, which are similar as two problem settings discussed earlier. Meanwhile, 334 testing data for GuessPassword are correctly predicted which represents 76.430% out of total 437. The same scenario as in the previous problem setting is observed for standard deviation value as SnmpGetAttack produces the lowest standard deviation value of 20.948.

Figure 3 shows the average rate obtained by each problem setting in 20 runs for all three types of attack investigated in this study. Average values shown in Tables 1-3 represents the average number of testing data that were correctly predicted in 20 runs. Average rate shown in Figure 3 determines the percentage of the average as compared to the total testing data for each attack type. SnmpGetAttack has the best average prediction accuracy (96.654%) when the proposed IGA employs 0.5 selection probability.

The same scenario is observed for GuessPassword attack type with average rate of 50.618%. Whereas for MailBomb, it produces the best average rate when 0.2 selection probability utilized.

For standard deviation analysis which calculate how much the obtained results deviate from the calculated mean value, the results for experiments with three different probability values are presented in Figure 4. From the results shown, all attack type produces the lowest standard deviation value using three different probability rates. These can be observed when SnmpGetAttack, GuessPassword and MailBomb have the smallest standard deviation when 0.3, 0.5, and 0.2 rate was employed respectively.

Table 3. Results for Experiments with 0.5 Probability of Random Data Selection

Type of Attack	Number of Data	The Best Result	The Worst Result	Average	Standard Deviation	Variance
SnmpGetAttack	774	774 (100%)	698 (90.181 %)	748.1 (96.654 %)	20.948	438.819
GuessPassword	437	334 (76.430 %)	134 (30.663 %)	221.2 (50.618 %)	60.722	3687.161
MailBomb	500	500 (100%)	384 (76.8 %)	460.95 (92.19 %)	47.529	2259.006

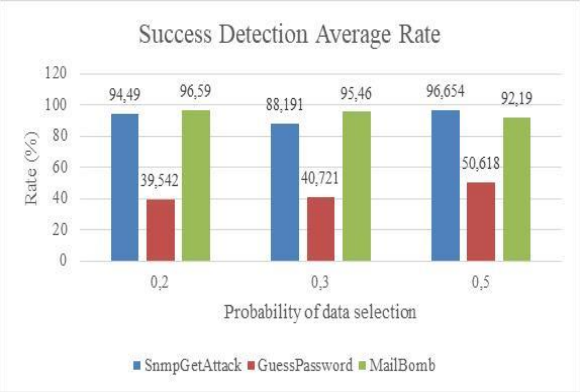

Figure 3. Average success rate for three different selection probabilities

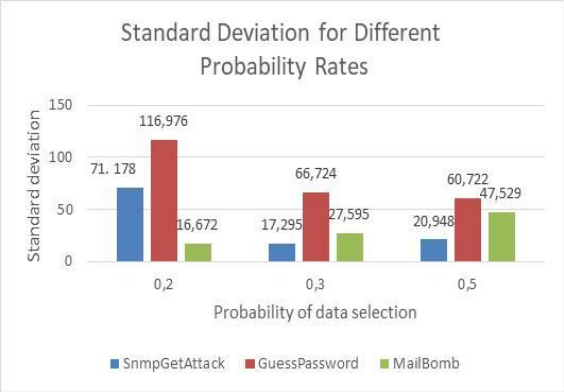

Figure 4. Standard deviation for experiments with different selection probability rates

4. CONCLUSION

In this study, Immune Genetic Algorithm (IGA) was utilized to predict the occurrence of intrusion in a computer network. Probability of data selection plays an important role to determine the accuracy of the prediction. Results obtained were analyzed using statistical analysis to investigate the performance of the proposed method. Fitness value evaluation during the training process determines the quality of candidate solution that can detect the occurrence of network attack. Based on the presented results, the proposed method has managed to correctly predict all testing data for SnmpGetAttack and MailBomb attack types. Whereas for GuessPassword, IGA was able to predicts the occurrence of intrusion for 372 out of 437 testing data. This represents 85% of prediction accuracy. Modification on the proposed method is warranted to improve its performance so that the accuracy of the prediction for GuessPassword attack type can at least reach the acceptable accuracy level of 95% as indicated in many previous studies.

ACKNOWLEDGEMENTS

The authors would like to express the gratitude to the Ministry of Education, Malaysia and Universiti Teknologi MARA, Selangor, Malaysia for the financial support given for this project (Geran Bestari) [600-IRMI/PERDANA 5/3 BESTARI (044/2018)].

REFERENCES

[1] A. H. Hamamoto, L. F. Carvalho, L. D. H. Sampaio, T. Abrão and M.L. Proença Jr, "Network Anomaly Detection System using Genetic Algorithm and Fuzzy Logic," *Expert Systems with Applications, 92,* pp. 390-402, Feb 2018.

[2] S. Mohammadi, *et al.,* "Cyber Intrusion Detection by Combined Feature Selection Algorithm," *Journal of information security and applications, 44,* pp. 80-88. Feb 2019.
[3] A. H. Hamamoto, L. F. Carvalho, L. D. H. Sampaio, T. Abrão and M.L. Proença Jr, "Network Anomaly Detection System using Genetic Algorithm and Fuzzy Logic," *Expert Systems with Applications, 92,* pp. 390-402, Feb 2018.
[4] S. Mohammadi, *et al.,* "Cyber Intrusion Detection by Combined Feature Selection Algorithm," *Journal of information security and applications, 44,* pp. 80-88. Feb 2019.
[5] N. T. Hanh, H. T. T. Binh, N. X. Hoai and M. S. Palaniswami, "An Efficient Genetic Algorithm for Maximizing Area Coverage in Wireless Sensor Networks," *Information Sciences,* Feb 2019.
[6] D.T.H. Ly, N. T. Hanh, H. T. T. Binh and N. D. Nghia, "An Improved Genetic Algorithm for Maximizing Area Coverage in Wireless Sensor Networks," *In Proceedings of the Sixth International Symposium on Information and Communication Technology,* pp. 61-66, Dec 2015.
[7] D. Jianjian, T. Yang and Y. Feiyue, "A Novel Intrusion Detection System based on IABRBFSVM for Wireless Sensor Networks," *Procedia computer science, 131,* pp. 1113-1121, Dec 2018.
[8] A. Shenfield, D. Day and A. Ayesh, "Intelligent Intrusion Detection Systems using Artificial Neural Networks," *ICT Express, 4*(2), pp. 95-99, Jun 2018.
[9] S. Roshan, *et al.,* "Adaptive and Online Network Intrusion Detection System using Clustering and Extreme Learning Machines," *Journal of the Franklin Institute, 355*(4), pp. 1752-1779, March 2018.
[10] A. Javaid, *et al.,* "A Deep Learning Approach for Network Intrusion Detection System," *In Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS),* pp. 21-26, May 2016.
[11] V. Hajisalem, and S. Babaie, "A Hybrid Intrusion Detection System Based on ABC-AFS Algorithm for Misuse and Anomaly Detection," *Computer Networks, 136,* pp. 37-50, May 2018.
[12] N. Hubballi and V. Suryanarayanan, "False Alarm Minimization Techniques in Signature-Based Intrusion Detection Systems: A Survey," *Computer Communications, 49,* pp. 1-17, August 2014.
[13] B. Narendra Kumar, M. Sivarama Bhadri Raju and B. Vishnu Vardhan, "A Novel Approach for Selective Feature Mechanism for Two-Phase Intrusion Detection System," *Indonesian Journal of Electrical Engineering and Computer Science,* 14(1), p. 101, April 2019.
[14] J. Jabez and B. Muthukumar, "Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection Approach," *Procedia Computer Science, 48,* pp. 338-346, 2015.
[15] A. A. Amira, A. E. Hassanien and A. S. Mostafa, "Artificial Immune System Inspired Intrusion Detection System using Genetic Algorithm". *Informatica*. 36. pp. 347-357, Jan 2012.
[16] A. A. Amira, A. S. Mostafa, A. E. Hassanien and E. H. Sanaa, "Detectors Generation using Genetic Algorithm for a Negative Selection Inspired Anomaly Network Intrusion Detection System", *In proceeding of: IEEE FedCSIS, At Wroclaw, Poland,* pp. 625-631, 2012.
[17] M, Ole and E. G. David, "The Crowding Approach to Niching in Genetic Algorithms, Evolutionary Computation", *MIT Press Cambridge*, Vol. 16(3), pp. 315-354, 2008.
[18] A. Firas and N. Reyadh, "Fitness Function for Genetic Algorithm used in Intrusion Detection System". *International Journal of Applied Science and Technology*, Vol. 2, No. 4, pp. 129-134. April 2012.
[19] S. I. Suliman, M. S. A. Shukor, M. Kassim, R. Mohamad, and S. Shahbudin. "Network Intrusion Detection System using Artificial Immune System (AIS)." In *2018 3rd International Conference on Computer and Communication Systems (ICCCS),* pp. 178-182. IEEE, 2018. Detection.
[20] J. Jabez and B. Muthukumar, "Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection Approach," *Procedia Computer Science, 48,* pp. 338-346, Jan 2015.
[21] B. Selvakumar and K. Muneeswaran, "Firefly Algorithm Based Feature Selection for Network Intrusion Detection," *Computers & Security, 81,* pp. 148-155, March 2019.
[22] S. Aljawarneh, M. Aldwairi and M. B. Yassein, "Anomaly-Based Intrusion Detection System Through Feature Selection Analysis and Building Hybrid Efficient Model," *Journal of Computational Science, 25,* pp. 152-160, March 2018.
[23] T. Dash, "A Study on Intrusion Detection using Neural Networks Trained with Evolutionary Algorithms," *Soft Computing, 21*(10), pp. 2687-2700, May 2017.
[24] S. Elsayed, R. Sarker and J. Slay, "Evaluating the Performance of a Differential Evolution Algorithm in Anomaly Detection. In *2015 IEEE Congress on Evolutionary Computation (CEC),* pp. 2490-2497, May 2015.
[25] S. Das and P. N. Suganthan, "Differential Evolution: A survey of the State-of-the-Art," *IEEE Transactions on Evolutionary Computation,* vol. 15, no. 1, pp. 4–31, 2011.
[26] D. Vasumathi and T. S, "Scalarizing Functions in Solving Multi-Objective Problem-an Evolutionary Approach," *Indonesian Journal of Electrical Engineering and Computer Science,* 13(3), p. 974, March 2019.
[27] S. C. M. Nasir, M. H. Mansor, I. Musirin, M. M. Othman, T. M. Kuan, K. Kamil and M. N. Abdullah, "Multistage Artificial Immune System for Static VAR Compensator Planning," *Indonesian Journal of Electrical Engineering and Computer Science, 14*(1), pp. 346-352, April 2019.