Immune-Genetic Algorithm(IGA) with Local Search For Intrusion Detection System in Computer Network

Hamizan Suhaimi, Saiful Izwan Suliman, and Ismail Musirin

Abstract—**The Internet provides almost unlimited connectivity to the online world that is widely used in our daily lives nowadays. As for borderless connections, inventors have to face great challenges in providing the greatest quality of service specifically in terms of security. Even with existing security measures such as firewalls, Intrusion Detection System (IDS) and antivirus to protect the network, the network is still vulnerable and its resources can be compromised by third parties. This problem highlights the need to address network intrusion problems efficiently. By formulating a specific algorithm for this problem, the purpose of this study is to examine the performance of improvised Genetic Algorithms for network intrusion problems. Based on the 1999 KDD Cup data set with various disruptions simulated from the military network, this research was conducted based on this standard dataset. The performance in terms of average intrusion detection rate and false alarm rate of the proposed method and other available techniques were analyzed to evaluate and determine the best performance. The combination of Genetic Algorithm, Immune Algorithm and local search has produced good detection with acuracy rate of 98.91% and has the potential to be further investigated for other research areas.**

Index Terms— **Intrusion Detection System, Genetic Algorithm, computer network systems.**

I. INTRODUCTION

IN this era of borderless world, most companies that provide communication services compete with each other to provide the best services in terms of data transfer rates including extensive internet coverage [1]. The need of Intrusion Detection System (IDS) is very crucial as there is a high risk that any device in the network could be compromised by a third party for unethical purposes. IDS works by monitoring network traffic to detect malicious connection for software or physical applications such as traffic that violates security and acceptable policies. There are two categories of IDS technologies which are Signature based IDS and Anomaly based IDS for variance of network configurations [2][3]. Each has different disadvantages and advantages in terms of configuration, detection and cost.

For example, an attacker can change traffic but traffic cannot be detected by IDS based on signatures compared to IDS based on anomalies. In this case, signature based IDS is only capable in analyzing unwanted traffic from known traffic.

Intrusion Detection System (IDS) is designed to detect any intrusion by monitoring network traffic and detect any suspicious activity such as illegal network access and malicious attacks [4]. Previous research has shown that they do not consistently obtain detection rates higher than 90% [2]. This can be due to the selection of data for training and testing process which was done based on probability. In addition, the main source of the raw data contains some errors that can increase false negative rate. This study was conducted to evaluate the performance of different methods used in tackling such problem. It focuses on the effect of mutation and data selection probabilities on detection rate in each iteration. The proposed method utilized in this study is Genetic Algorithm combined with a crucial step in Artificial Immune System AIS) to improve the generated candidature solutions. This algorithm is further enhanced with the utilization of taboo tenure extracted from Tabu Search algorithm, hence the name of the proposed method.

II. PAST RESEARCH

There are many researchers who have proposed GA with additional method(s) to make improvements to the proposed method. An improvement method for GA has been proposed by a group of researchers from the University of Mumbai. P. U. Kadam and P. P. Jadhav [7]. In their research, GA was utilized with all forty-one features for the chromosomes. Using a newly developed solution, it has been evaluated using a fitness function to calculate the decency of each chromosome based on the desired solution.

B. Chakrabarty et. al has conducted studies on IDS based on anomalies using GA hybrids and K-Centroid Clustering [2]. In this study, two datasets which are KDD'99 Cup and NSLKDD were used to conduct experiments for IDS.NSLKDD data set has been identified to be able to solve problems in the KDD'99 data set. For this research, GA with K-Centroid Clustering was performed on the specific clusters. Clusters were created to differentiate groups for specific disruptive and normal data. During clustering, different clusters were also created for

This manuscript is submitted on 21th August 2020 and accepted on 2th February 2021. This work was supported by Universiti Teknologi MARA.

Hamizan Suhaimi, Saiful Izwan Suliman and Ismail Musirin are with the Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Malaysia (e-mail: saifulizwan@uitm.edu.my).

similar connection types. S. Sharma et. al has make a review of GA and Fuzzy Logic approaches for IDS [8]. Since KDD'99 dataset has some redundancy issues, NSLKDD was used as the main dataset. As a result of this study, the fuzzy-GA hybrid has performed better than the single GA technique.

GA with new features selection has been proposed with SVM classification to improve detection accuracy [9]. Data collectors and pre-processing modules have been proposed for data pre-processing, extraction and classification of data. This pre-processing technique is used for data reduction from the KDD99 data set where only the most relevant features are selected and used for detection. According to [10], GA is proposed together with SVM to overcome the problem of difficulty in identifying new data and the need for large data sets to do so. The proposed GA with SVM has resulted in a better detection rate with a low false positive rate compared to the Advanced Feature Selection Algorithm (FFSA) and Linear Correlation Feature Selection (LCFS).

A study for IDS using hybrid GA-SVM and selection of new features has been proposed by H. Gharaee and Hosseinvand [11]. In their research, a new fitness function has been introduced with the Least Squares Support Vector Machine (LSSVM). Using the KDD'99 and UNSW-NB15 Cup data sets, the proposed method has involved three main steps: character selection based on GA, training and classification. As a first step of detection, LSSVM is trained in the training division. Then, traffic data is classified into normal or anomalous classes. In this paper, three parameters have been introduced for the new fitness function which are Positive Rate (FPR), True Positive Rate (TPR) and the number of features selected to calculate each subset of features. In producing results, MMIFS model using KDD'99 Cup has been applied into LSSVM model to compare with other methods such as GA-Fuzzy SVM, MMIF, SVM and C4.5.

Feature selection of GA with multiclass SVM classifier was proposed by Ganapathy et al [12]. In their research, only 10 features from 40 features were selected for the new dataset. The proposed technique has managed to get better accuracy compared to other techniques. In [13], SVM classifier was proposed for GA to select the most importance features and useless elements in recognizing attacks. Instead of using KDD Cup 99, NSL-KDD is used as the main dataset using only 28 features while the rest features have been omitted. In 2016, A. S Desai and D. P Gaikwad proposed the Fuzzy-Genetic Algorithm hybrid method to identify internal and external attacks of network systems [14]. Signature matching algorithms can be used to identify attacks from the internal system while FGA was proposed to detect external attacks in the network. In the internal systems, SQL injection which consists of static and dynamic injection detection became the main focus. In static parts, query are classified as malicious attack if the comparison is match with the stored signature, otherwise it is normal. In dynamic detection, query are considered as malicious if it exceeds the threshold value when using calculation of similarity index from the comparison.

Studies on the combination of 2 techniques have been done over the past few years. The combination of Artificial Neural Networks (ANN) and GA for IDS was done by A. Dastanpour et al. to study the effectiveness of both techniques [15]. A comparison was also made between the proposed technique with the GA-SVM model to evaluate and observe the performance of both models. Both models function as classifiers for GA, used for the process of recognition and data classification into groups according to attack and normal traffic. Despite only using a low number of features, GA-ANN (18 features) managed to produce good performance and more effective than GA-SVM which uses 24 features. Based on the study by the same reseachers [16], GA technique generates the first feature at random while the remaining features were formed by GA and ANN classification. In this research, a comparison of achievements for the proposed GA-ANN was conducted with other techniques such as Modified Mutual Information-based Feature Selection (MMIFS), Forward Feature Selection Algorithm (FFSA) and Linear Correlation Feature Selection (LCFS). The results of the study found that GA-ANN can achieve the highest detection rate using only 18 features while other techniques require at least 21, 24 and 34 features respectively to get the same result. Feature selection for proposed GA and ANN classification acts as the main cause to produce good results even if the data features are reduced to a minimum value.

According to a study from J. Kaliappan et al. [17], ANN can be used to derive meaning, analyze and process information from complex noisy data. Supervised learning is the process of ANN being able to learn a task with random responses. Reactions can weaken the connection when the wrong decision is made or strengthen the connection for the right decision. Studies using K-Nearest Neighbor (KNN) and GA for IDS were also conducted by researchers to improve computer network security. According to Y. Canbay and S. Sagiroglu [18], KNN is generally used by test samples 'neighbors voting to determine the class labels from the test set. For hybrid GA, KNN is used to calculate the distance and to classify test samples by creating neighborhood diversity using its operators. After scanning all training data, KNN will calculate the distance between each test sample and training sample. KNN has several limitations such as high calculation complexity and training set dependency where it needs all training datasets to create a model [19]. Eventhough, KNN is an efficient algorithm and easy to implement [20].

III. THE PROPOSED METHOD

The proposed Immune-Genetic Algorithm (IGA) with local search involved 3 main phases which are data division, training and testing phase. The main steps involved in this IGA system are discussed in the following sections:

A. Data division

The dataset utilized in this study was taken from KDD-99 The Fifth International Conference on Knowledge Discovery and Data Mining [21]. A total of 251,411 raw connection data were investigated in this study. The first step taken before the implementation of the proposed algorithm is the process of separating the main data set into two groups. During this phase, raw data were divided according to the standard formulations of 10% and 90% for the respective testing and training process. Normal and intrusion connections were chosen at random based on the probabilities value for the algorithm. Preliminary investigations were performed with probabilities of 0.2, 0.3 and

0.5 for the data separation process using IGA method. This process is important to determine the optimal value that will result in the best average of intrusion detection. In this process, each probability value was executed twenty times and the results were recorded and analyzed. Probability value of 0.3 was finally chosen to be applied in the IGA-Local Search and other method (GA and AIS) as it produces the best average of detection rate.

B. Training process

The training process for the proposed IGA with local search is shown in **Error! Reference source not found.**. The initial population was randomly generated and then trained using Immune-Genetic Algorithm (IGA) along with local search improvement methods. This training process involves processes such as data selection, crossover (recombination), mutation, and data regeneration. In the process of chromosome selection, only the top 20 individuals with the highest fitness value were selected to be cloned twice and subsequently crossover process occurs. In this implementation, crossover rate was used to perform chromosome regrouping. Based on the standard crossover process, the probability is used to determine how often a crossover will be performed in training process. In the proposed

algorithm, a child will be produced each time the crossover rate is met. This chromosome is then mutated based on the mutation probability set at 0.15 as it produces the best average detection compared to other probabilities tested. Fitness value for the mutated chromosome is calculated. The new population is then sorted from the best to worst.

Step 1: Generate initial population of 100 chromosomes.
Step 2: Measure fitness value.
Step 3: Data sorted in ascending order based on fitness value.
Step 4: Select the 20 worst chromosomes and store in tabu file.
Step 5: Select top 20 good-quality chromosomes.
Step 6: Clones 2 times each of the 20 chromosomes.
Step 7: Crossover between 2 parents of chromosomes.
Step 8: Mutate the chromosomes.
Step 9: Calculate fitness value.
Step 10: Sort the chromosomes in ascending order based on the fitness value. .
Step 11: Take top 30 chromosomes from mutated data, 30 intial chromosomes and 40 new chromosomes generate randomly to form a new population.
Step 12: Apply tabu search for filtering process.
Step 13: Repeat 100 times of attack recognition between the population and training data.
Step 14:Final population to compare with testing data.

Figure 1: Training process using IGA-Local Search in IDS

A new population of chromosomes is then produced from a combination of the 30 best chromosomes from the initial population, 30 mutated chromosomes and 40 newly produced chromosomes. The fitness value of the chromosomes will be calculated, and the chromosomes will undergo the same training process for the next 90 iterations.

After the 90th iteration, a tabu list is constructed. This list consists of twenty chromosomes generated for the initial solution. In the last ten iterations, each new chromosome will be compared with the chromosomes stored in the tabu list. If it is of worst quality, a new chromosome will be regenerated. This process is repeated until all the chromosomes produced in the population have better fitness value than the chromosomes stored in the tabu list. This process is performed based on the concept of tabu search, which is considered as one of the effective local search method. It is done with the hope of generating good-quality solutions for this intrusion detection problem.

Step 1: Save 100 population of chromosomes from training
process into new file.
Step 2: Detect intrusion using 100 trained chromosomes and
chromosomes of testing data.
Step 3: Save the total number of unrecognized attacks into "Unrecognized Testing Data.txt".
Step 4: Display the total of successful detection and
false alarm value

Figure 2: A Testing process of IDS

C. Testing process

Figure 2 above shows the testing process where the final population from the training process was used to identify whether data in testing file is an attack or not. Testing data which is successfully recognized is considered as success detection either it is correctly predicted as normal/attack connection or vice versa. During the testing phase, a total of 24920 connections were investigated. These connections refer to a set of internet connections with specific attributes which were extracted from the main dataset (KDD Cup'99).

In order to evaluate the performance of the proposed method during the testing phase, 4 categories of detection were identified which are true positive, false negative, true negative and false positive [5][6]. All categories are calculated using the equations below:

$$True\ Positive\ Rate\ (TPR) = \frac{TP}{TP+FN} x100\% \qquad (1)$$

$$False\ Negative\ Rate\ (FNR) = \frac{FN}{TP+FN} x100\% \qquad (2)$$

$$True\ Negative\ Rate\ (TNR) = \frac{TN}{TN+FP} x100\% \qquad (3)$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{TN+FP} x100\% \qquad (4)$$

$$Success\ Detection\ Rate = \frac{TP+TN}{Total\ testing\ data} x100\% \quad (5)$$

Equation **Error! Reference source not found.**) identifies the true positive rate which is calculated based on the number of testing data that is correctly predicted as an attack connection. The false negative rate which is calculated using equation **Error! Reference source not found.**) identifies the number of attack connection (from the testing data) which are predicted as normal connection. Equation **Error! Reference source not found.**) will calculate the true negative rate (TNR) based on number of normal connection that is correctly predicted whereas false positive rate (FPR) identifies the number of normal connection that is predicted as attack connection. The effectiveness of the proposed algorithm can be evaluated by using equation **Error! Reference source not found.**) based on the number the testing data that is correctly predicted.

IV. Results

The proposed IGA with local search has been implemented on IDS using data selection and mutation probabilities of 0.3 and 0.15 probability respectively. The values were set based on the test of the effectiveness of the probability value of IGA on the IDS simulation. Both probabilities produced high average attack detection rate compared to other probability values. The result can be seen in Figure 3.

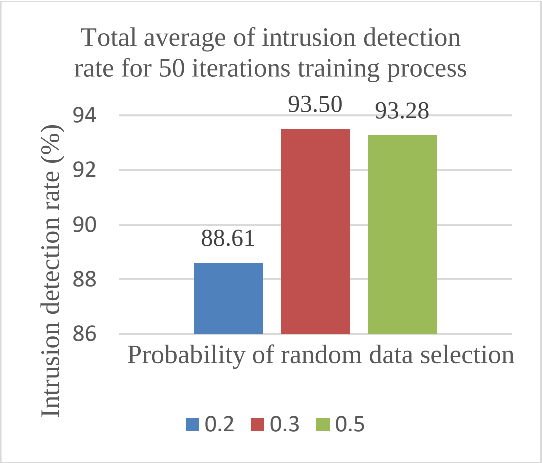

Figure 3: Results for 3 different selection probabilities on IDS in 50 iterations

Three probabilities of initial data selection were investigated to evaluate their performance on this IDS and the results are shown in Figure 3. Based on the investigation, probability of 0.3 produces the best detection rate among the three with 93.504% accuracy of detection. This is followed by 0.5 and 0.2 probabilities with 93.276% and 88.614% accuracy respectively.

This experiment were conducted by using 0.15 probability of mutation as it produces the best result from the previous investigations. The training process was conducted in 50 iterations before the testing phase took place.

Figure 4 shows the results for the same problem setting as utilized in Figure 3 except that the training process was conducted in 100 iterations. The same pattern of results can be observed as presented in Figure 3. Probability of initial data selection of 0.3 produces the best detection rate with 94.69% accuracy. Meanwhile, probability of 0.5 produces 92.855% accuracy and followed by 0.2 with 91.503%.

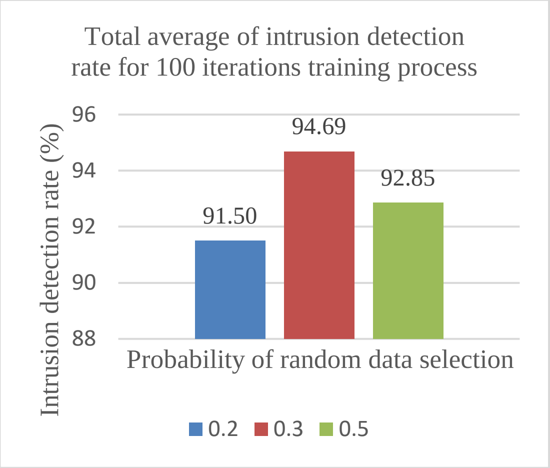

Figure 4: Results for 3 different selection probabilities on IDS in 100 iterations.

Based on the results shown in Figure 3 and Figure 4, probability of initial data selection of 0.3 was chosen to be utilized in our proposed algorithm and the results is presented in Figure 5.

The bar chart illustrated in Figure 5 shows the average of intrusion detection rate for all five types of attack by using the proposed IGA-LS. These average values were obtained after a series of twenty runs by using probability of initial selection of 0.3 and probability of mutation of 0.15. The proposed method performs the best in detecting Smurf type of attack with 99.993% accuracy of detection. This is followed by Mailbomb (99.87%) , SnmpGetAttack (98.404%) and Neptune (97.972%). These results indicate the efficiency of IGA-LS on IDS with high detection rate (more than 97%). However, the proposed method is having difficulties in detecting GuessPassword as it only manage to produce average of 64.989% accuracy in 20 runs.

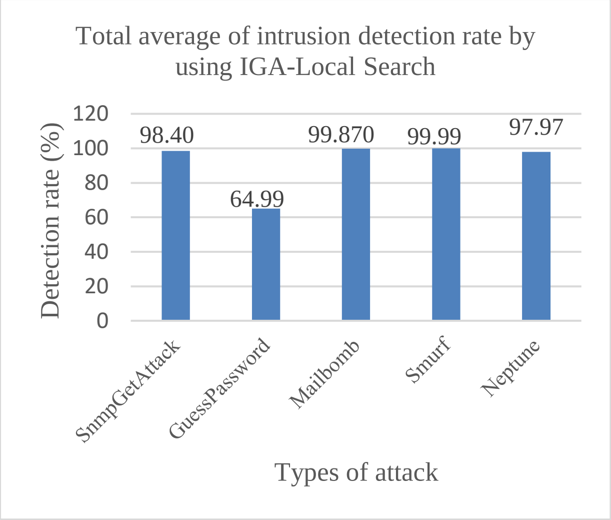

Figure 5: Total average of intrusion detection rate for each attack based on 0.3 probabilities of data selection

TABLE 1
DETECTION RESULTS FOR ALL IDS TECHNIQUE.

Type of Attack	Immune-GA	Immune-GA with local search	Standard GA	AIS
SnmpGetAttack	682.60 (88.19 %)	761.65 (98.40%)	751.75 (97.12%)	731.25 (94.47%)
GuessPassword	177.95 (40.72 %)	284 (64.98%)	119.35 (27.31%)	145.90 (33.38%)
Mailbomb	477.30 (95.46 %)	499.35 (99.87%)	433.40 (86.68%)	483.85 (96.77%)
Smurf	16217.35 (98.83%)	16407.80 (99.99%)	16398.05 (99.93%)	16400.75 (99.95%)
Neptune	5094.70 (87.84%)	5682.35 (97.97%)	4770.40 (82.24%)	4632.15 (79.86%)
Total Average	94.69%	98.80%	93.95%	93.62%

In order to evaluate the performance of the proposed IGA-LS, comparative studies with other methods were conducted on the same dataset. A standard Genetic Algorithm (GA), Artficial Immune System (AIS) and Immune-Genetic Algorithm (IGA) were chosen for this purpose. The details comparison are shown in **Error! Reference source not found.**. Within 20 series of run, IGA-LS produces the highest attack detection rates for all five attacks investigated compared to other 3 methods. Immune-GA is the second best method for GuessPassword and Neptune attack with an average of 40.721% and 87.84% accuracy rate respectively. AIS algorithm is superior than standard GA and Immune-GA for Mailbomb and Smurf attack. Meanwhile, standard GA is better in detecting SnmpGetAttack with 97.125% detection average rate. In overall, IGA-LS produces 98.809% positive detection rate for all five attacks. This is followed by Immune-GA (94.69%), standard GA (93.95%) and AIS (93.62%). These results are illustrated in Figure 6. From the results shown in Table 1, GuessPassword is the most difficult attack to be detected. Immune-GA with local search is the best performer with only 65% of the testing data accurately identified. This is because of the properties for the GuessPassword attacks vary distinctively one another. Hence, making it difficult for the algorithms to correctly identify the patterns.

The dataset utilized in this study consists of two types of connection data, which are attack connection and normal connection. The results obtained were further analyzed based on these two categories.

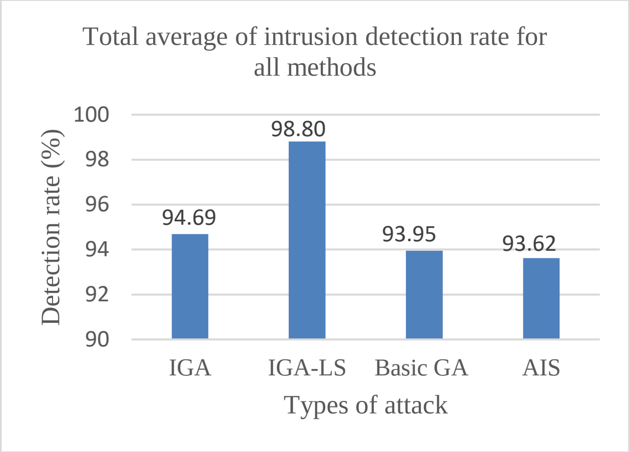

Figure 6: Total average of intrusion detection rate for 4 different techniques of IDS

As discussed in Section II, four detection categories were investigated which are True Positive Rate (TPR), False Negative Rate (FNR), True Negative Rate (TNR), and False Positive Rate (FPR). The values for these four rates were calculated and shown in

for all methods. For attack connection detection in testing phase, 23920 testing data were utilized.

The proposed Immune-Genetic Algorithm with local search (IGA-LS) method produces the highest TPR value of 99.639% as compared to other three methods. It shows that IGA-LS has managed to identify the most number of attack connection. This is also reflected in False Negative Rate with IGA-LS produces the lowest value which indicates the least number of attack connection wrongly detected. However, this is not the case for normal connection detection. The proposed IGA-LS is the worst-performing method with the lowest True Negative Rate (5.7%). AIS is the best for this criteria with 30.5% successful detection rate. When both attack and normal connections are considered, IGA-LS produces the best result with 95.41% of successful detection rate. This is followed by IGA with 95.34%. Even though the difference is quite small which is 0.07%, it shows that local search has managed to slightly increase the detection rate. However, this small value represents more than 1700 testing data which is very crucial in making sure a computer network is secured from any illegal intrusion.

TABLE II
DETECTION RATES FOR ALL METHODS BASED ON DIFFERENT PARAMETERS

Types of Detection Rate		**IGA (%)**	**IGA-LS (%)**	**Standard GA (%)**	**AIS (%)**
Condition Positive (Attack)	True Positive Rate (TPR)	99.19	99.63	99.27	99.24
	False Negative Rate (FNR)	0.81	0.361	0.726	0.761
Condition Negative (Normal)	True Negative Rate (TNR)	19.30	5.70	16.40	30.50
	False Positive Rate (FPR)	80.70	94.30	83.60	69.50
Total Detection (Success Detection Rate)		95.34	95.41	94.51	84.43

V. Conclusion

There are many aspects that can influence the effectiveness of a network intrusion detection system. For a population based metaheuristic like Genetic Algorithm, the initial value of all parameters involved is very crucial as it will determine the quality of the produced solutions. In this study, the probability of the initial data selection was investigated. This probability value will determine the type of connections that will be allocated in both training and testing dataset. This allocation process can be a decisive factor to make sure the population of antibodies (a set of candidate solutions) is good enough to be used to detect any occurrence of network intrusion. The combination of optimal probability values for initial data selection and mutation on the proposed IGA-LS helps in improving the average intrusion detection rate as well as True Posite Rate.

The utilization of a local search method on the Immune-Genetic Algorithm (IGA) is aimed to improve the Positive Detection Rate. This was investigated in this study as the framework for tabu search method was modelled and implemented in solving this intrusion detection issue. The presented results have proven that the proposed local search has the capability to further improve the detection rate, thus showing its effectiveness in helping IGA to explore the solution neighbourhood for better outcomes.

Acknowledgment

The authors also would like to express the gratitude to the Ministry of Education, Malaysia for the financial support given for this project (Geran Bestari) [600-IRMI/PERDANA 5/3 BESTARI (044/2018)]. A special appreciation to Universiti Teknologi MARA, Selangor, Malaysia for providing facilities such as laboratory and library that were used to finish this research.

References

[1] J. Soni, “Machine Learning based approach for solving Intrusion Detection System,” vol. 5, no. 5, pp. 1–6, 2016.

[2] B. Chakrabarty, “Anomaly based Intrusion Detection System using Genetic Algorithm and K - Centroid Clustering,” *Int. J. Comput. Appl.*, vol. 163, no. 11, pp. 975–8887, 2017.

[3] P. S. Bhattacharjee, “Intrusion Detection System for NSL-KDD Data Set using Vectorised Fitness Function in Genetic Algorithm,” *Adv. Comput. Sci. Technol.*, vol. 10, no. 2, pp. 235–246, 2017.

[4] S. I. Suliman, M. S. Abd Shukor, M. Kassim, R. Mohamad, and S. Shahbudin, “Network Intrusion Detection System Using Artificial Immune System (AIS),” *2018 3rd Int. Conf. Comput. Commun. Syst. ICCCS 2018*, pp. 426–430, 2018.

[5] K. Kumar and J. Singh, “Network Intrusion Detection with Feature Selection Techniques using Machine-Learning Algorithms,” *Int. J. Comput. Appl.*, vol. 150, no. 12, pp. 1–13, 2016.

[6] P. Tao, Z. Sun, and Z. Sun, “An Improved Intrusion Detection Algorithm Based on GA and SVM,” *IEEE Access*, vol. 6. pp. 13624–13631, 2018.

[7] P. U. Kadam and P. P. Jadhav, “An effective rule generation for Intrusion Detection System using Genetics Algorithm,” vol. 2, no. 10, 2014.

[8] S. Sharma, S. Kumar, and M. Kaur, “Recent trend in Intrusion detection using Fuzzy- Genetic algorithm,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 3, no. 5, May 2014, pp. 6472–6476, 2015.

[9] B. Senthilnayaki, D. K. Venkatalakshmi, and D. A. Kannan, “Intrusion Detection Using Optimal Genetic Feature Selection and SVM based Classifier,” *2015 3rd Int. Conf. Signal Process. Commun. Netw.*, pp. 1–4, 2015.

[10] A. Dastanpour and R. A. R. Mahmood, “Feature selection based on genetic algorithm and SupportVector machine for intrusion detection system,” *Int. Conf. Informatics Eng. Inf. Sci.*, no. September 2014, pp. 169–181, 2013.

[11] H. Gharaee and H. Hosseinvand, “A new feature selection IDS based on genetic algorithm and SVM,” in *2016 8th International Symposium on Telecommunications, IST 2016*, 2017, pp. 139–144.

[12] S. Ganapathy, P. Yogesh, and A. Kannan, “Intelligent agent-based intrusion detection system using enhanced multiclass SVM,” *Comput. Intell. Neurosci.*, vol. 2012, 2012.

[13] B. M. Jahromy, “A New Method for Network Intrusion by Using a Combination of Genetic Algorithm and support Vector Machine Classier.” pp. 1–9.

[14] A. S. Desai and D. P. Gaikwad, “Real time hybrid intrusion detection system using signature matching algorithm and fuzzy-GA,” *2016 IEEE Int. Conf. Adv. Electron. Commun. Comput. Technol. ICAECCT 2016*, pp. 291–294, 2017.

[15] A. Dastanpour, S. Ibrahim, R. Mashinchi, and A. Selamat, “Comparison of genetic algorithm optimization on artificial neural network and support vector machine in intrusion detection system,” *ICOS 2014 - 2014 IEEE Conf. Open Syst.*, pp. 72–77, 2014.

[16] A. Dastanpour, S. Ibrahim, and R. Mashinchi, “Using Genetic Algorithm to Supporting Artificial Neural Network for Intrusion Detection System,” *Int. Conf. Comput. Secur. Digit. Investig.*, pp. 1–13, 2014.

[17] J. Kaliappan, T. Revathi, and S. Karpagam, “Intrusion Detection using Artificial Neural Networks with Best Set of Features,” no. October, 2015.

[18] Y. Canbay and S. Sagiroglu, “A Hybrid Method for Intrusion Detection,” *2015 IEEE 14th Int. Conf. Mach. Learn. Appl.*, pp. 156–161, 2015.

[19] Y. Chang, “Semi-supervised Classification Algorithm Based on the KNN,” pp. 9–12, 2011.

[20] Y. Tan, “An Improved KNN Text Classification Algorithm Based on K-Medoids and Rough Set,” *Proc. - 2018 10th Int. Conf. Intell. Human-Machine Syst. Cybern. IHMSC 2018*, vol. 1, no. 1, pp. 109–113, 2018.

[21] https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

S. Hamizan is a máster student by research at the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Malaysia. He graduated in Diploma of Electrical Engineering (Electronic) in 2014 and obtained 1st class bachelor degree (Honor) in Electronic Engineering from UiTM in 2018. His research interest is mainly in the area of computer network especially genetic algorithm and intrusion detection system.

.

Saiful I. Suliman is a senior lecturer at the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Malaysia. He obtained 1st class degree in Artificial Intelligence in 2002 and Master of Science (Electrical Engineering) from UiTM in 2006. His Phd was awarded by The University of Nottingham, UK in 2015. His research interest is mainly in the area of artificial intelligence, optimization and pattern recognition algorithm, metaheuristic approach, frequency bandwidth spectrum and power system operations. He has published papers in many international conferences as well as reputable journals. He is a certified *Professional Technologist* and *Chartered Engineer* (*CEng*) with IET, UK. He is currently the head of Innovation Unit, Research Management Centre (RMC) at UiTM.

Ismail Musirin received the bachelor’s degree (Hons.) in electrical engineering from Universiti Teknologi Malaysia, in 1990, the M.Sc. degree in pulsed power technology from the University of Strathclyde, U.K., in 1992, the Ph.D. degree in electrical engineering from Universiti Teknologi MARA (UiTM), Malaysia, in 2005, and the Diploma degree in electrical power engineering from Universiti Teknologi Malaysia (UTM), in 1987. He is currently a Professor of power system with the Faculty of Electrical Engineering. He has authored and coauthored two books, more than 300 articles in international journal and indexed conferences. His research interests include artificial intelligence, optimization techniques, power system analysis, renewable energy, distributed generation, and power system stability. He is a Senior Member of the International Association of Computer Science and Information Technology (IACSIT), and a member of the Artificial Immune System Society (ARTIST) and of the International Association of Engineers (IAENG). He is also an International Journal Reviewer of the IEEE Transactions, Science (Elsevier), WSEAS, John Wiley, IET, and some other publishers.