Journal of Computer Science 10 (6): 1015-1025, 2014
ISSN: 1549-3636

doi:10.3844/jcssp.2014.1015.1025 Published Online 10 (6) 2014 (http://www.thescipub.com/jcs.toc)

HYBRID FEATURE SELECTION ALGORITHM FOR INTRUSION DETECTION SYSTEM

Seyed Reza Hasani, Zulaiha Ali Othman and Seyed Mostafa Mousavi Kahaki

Department of Artificial Intelligence Technology, Faculty of Information Science and Technology, University Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia

Received 2013-11-20; Revised 2014-01-27; Accepted 2014-01-31

ABSTRACT

Network security is a serious global concern. Usefulness Intrusion Detection Systems (IDS) are increasing incredibly in Information Security research using Soft computing techniques. In the previous researches having irrelevant and redundant features are recognized causes of increasing the processing speed of evaluating the known intrusive patterns. In addition, an efficient feature selection method eliminates dimension of data and reduce redundancy and ambiguity caused by none important attributes. Therefore, feature selection methods are well-known methods to overcome this problem. There are various approaches being utilized in intrusion detections, they are able to perform their method and relatively they are achieved with some improvements. This work is based on the enhancement of the highest Detection Rate (DR) algorithm which is Linear Genetic Programming (LGP) reducing the False Alarm Rate (FAR) incorporates with Bees Algorithm. Finally, Support Vector Machine (SVM) is one of the best candidate solutions to settle IDSs problems. In this study four sample dataset containing 4000 random records are excluded randomly from this dataset for training and testing purposes. Experimental results show that the LGP_BA method improves the accuracy and efficiency compared with the previous related research and the feature subcategory offered by LGP_BA gives a superior representation of data.

Keywords: IDS, Linear Genetic Programming, Feature Selection, Bees Algorithm, Anomaly Detection

1. INTRODUCTION

The computing networks had become an absolute tool for various sectors which includes social, economies, military and so on. It ensures the connectivity, collaboration and cooperation between these different sectors. The sensational developments of networks are naturally accompanied by the increase in the number of users. This tremendous increase in computer network usage and high accessibility of the internet gained many positive aspects. On the other hand, the raise of computer hacking is become thoughtful issue. To avoid of getting anomaly intrusions in network some security tools such as firewalls, antiviruses, internet security tools and network Intrusion Detection System (IDS) methods are developed to guard the computer servers and clients in all around the world. Anonymous and identified users are not necessarily full of good intentions on these networks. They can exploit the vulnerabilities of networks and systems. They can also access to sensitive or confidential information in order to read, modify or destroy them. Therefore, the act for networks security has become more significant in order to secure the networks from becoming the target of potential attacks (Folorunso *et al*., 2010).

IDS has been classified into two categories which are signature based anomaly detection and the anomaly behavior detection. The differences of these two types are in their patterns. The signature base intrusions regularly examine the network and try according to some predefined patterns matches on the network. whereas the anomaly network intrusion based systems provide normal traffic patterns and try to find the similar packets to be detected as an intrusions (Trair *et al*., 2007).

Corresponding Author: Seyed Reza Hasani, Department of Artificial Intelligence Technology, Faculty of Information Science and Technology, University Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia

According to (Lazarevic *et al*., 2003) the key element of intrusion detection can be known as anomaly detection or undesirable persuaded defects, attacks, faults.

The main goal of IDS is to prevent the happening of intrusions in the following transactions in the network by classifying packets into two types of attacks and normal. One of the most important things in the IDS is computational speed and comparison accuracy. According to the tremendous features in each transaction a proper algorithm is required to derive an effective subset of features in order to recognize the intrusions as well as full feature set.

Feature extraction is the process of determining subset of features (M) from an original set (N) where M<N. the purpose of feature selection is to crop those features that have more valuable information and contain some data about some of the other features. Furthermore, eliminating some of the redundant or irrelevant features is reducing the process time with minimum of accuracy decreasing. To achieve this target, so far some data mining approaches are introduced and performed to do the feature extraction (Chen *et al*., 2006).

Wrapper based and filter based methods are to approaches for feature selection in IDS. Wrapper based method utilizes the machine learning approaches to access the reliable features in intrusion detection. Where machine learning is never been used by the filter method to reduce the redundant or irrelevant features to create a minimized and effective feature set (Chen *et al*., 2006).

1.1. Related Works

There are several data mining solutions for intrusion detection while it has the ability and helpful information which demonstrate a user's behavior from large dataset. For anomaly detection, data mining has been applied (Lunt *et al*., 1992; Lee and Stolfo, 2000). Statistics (Debar *et al*., 1992; Anderson *et al*., 1995), Artificial Neural Network (ANN) (Lippmann and Cunningham, 2000; Cho and Park, 2003), Support Vector Machines (SVM) (Premanode *et al*., 2013), Neuro-Fuzzy (NF) computing (Mukkamala *et al*., 2005; Nirmala and Gowri, 2013), Multivariate Adaptive Regression Splines (MARS) (Banzhaf *et al*., 1998) and Linear Genetic Programming (LGP) (Mukkamala *et al*., 2006), Rough Set (Guo *et al*., 2010), Rough-DPSO (Rahman *et al*., 2009), BA (Alomari and Othman, 2012). These methods are commonly applied for misuse and anomaly detections.

LGP has been formerly effectively performed to a diversity of machine learning problems. The article (Banzhaf *et al*., 1998) examines the competency of linear genetic programming methods, for making intrusion detection methods and recognizing the major aspects that help in resolving whether a relation is normal activity or intrusive normal activity. The writers also compare the efficiency of LGP with SVM and a Neural Network (NN) trained exerting resilient back propagation learning. Performance metrics consist a few crucial phases of intrusion detection such as testing times and training, detection accuracy and scalability that assist IDSs do in near real time or real time.

To rapid up the evaluation process, the LGP approach is applying the new generations for lower rate instances (Sequeira and Zaki, 2002). As introduced in (Banzhaf *et al*., 1998) a multiple measure intrusion detection technique is offered by (Hettich and Bay, 1999) to defeat the impediment of single-measure detectors. Hidden Markov Model (HMM), rule-based and statistical methods are combined using a rule-based method. Chebrolu *et al*., (2004) introduced an IDS method to intrusion detection which combined the Bayesian network and regression tree.

Another AI method for anomaly detection is to select those properties that different from the rest of data (Lazarevic *et al*., 2003). SVM is represented to train these selections by training vectors in multiple data behavior.

In NF computing the linguistic rules representing forms of knowledge which can create Fuzzy Inference Systems (FIS) (Abraham , 2001). This process can be considered as pre-processing of ANN to take advantage of learning capability of FIS (Chavan *et al*., 2004; Jaganathan *et al*., 2013). Multivariate Adaptive Regression Splines (MARS) is known as a method to avoid of having binary and related variables. This invention is detecting the elements of all complex data transformation to eliminate the high dimensional data (Banzhaf *et al*., 1998; Abraham and Steinberg, 2001).

1.2. Proposed Method

This section introduces a new wrapper based optimization method of LGP_BA based on Linear Genetic Programming (LGP) and Bees Algorithm (BA) to achieve an efficient feature selection algorithm. Firstly, modified LGP is used to generate firs candidate chromosome and then BA applies neighborhood search to perform the modification. Finally, SVM classifies the high order feature sets.

1.3. IDS Feature Selection Algorithm using LGP_BA Method

The feature selection in this research is result of combination of two optimization methods which are Linear Genetic Programming (LGP) and Bees Algorithm (BA), which named LGP_BA. This proposed method

(LGP_BA) used the wrapper method to provide a random search technique for feature deduction and uses (SVM) as the classifier. Many methods for enhancing the performances of LGP have been introduced so far (Baranidharan and Ghosh, 2012). In this work, LGP provides random selection of features among all features in dataset to perform new generations. In each generation the crossover and mutation will be applied to categorize the highest fitness values then an evolution method termed Queen-Bee Evolution (QBE) is introduced for LGP. QBE defined similarly to the nature in which the queen-bee, the fittest individual in a generation, crossbreeds with the other bees selected as parents by BA. This increases the exploitation of LGP. Moreover, its increases the probability which LGPs fall into precipitate convergence and outputs with a decrease in the LGP's performances. The fitness value for the features subset is comparative with its gained information. Lopez introduced a fitness function which constructed as follow (López-Pujalte *et al*., 2003): firstly, using the scalar products calculate the similarity index of the query vectors, then, sort the results in a decreasing order of similarity. Finally, calculates the fitness value (F) of the chromosome using Equation (1). In the computation of the information gain for only one feature according to the classes is proposed like the following:

$$F = \frac{1}{|D|}\sum_{i=1}^{|D|}\left(r(di)\sum_{j=1}^{|D|}\right) \quad (1)$$

where, $|D|$ is declared as the total number of chromosomes retrieved andr(d) is the function that returns the relevance of chromosomes.

To decrease the probability of premature convergence, some individuals in queen-bee evolution is strongly mutated. This reinforces the exploration of LGP. Queen-bee evolution has been tested with one combinational problem and two typical function optimization problems. It was shown from the experiments that queen-bee evolution could be a typical evolution method for enhancing the performances of LGP. **Figure 1** shows the how LGP and BA are combined to produce LGP_BA method.

The LGP_BA flowchart shows the dependencies of two methods to each other. The designed structure to implement this method is presented in the following pseudo code. The procedure starts with LGP and Modifies by BA. The required argument is declared in the **Fig. 2** and described in the subsection.

From the above algorithm flow, some differences between conventional LGP and LGP_BA can be found. Queen-bee evolution in comparison to normal evolution has two major differences, which are marked by asterisks in the algorithm. First, the population will be selected by m selection from n possibilities and the first stage fitness will be calculated. The main criteria in the LGP part is the fitness value must be more than 50. If the condition has been met by the fitness function the chromosome will be passed to the BA process otherwise LGP applies the crossover and mutation to create some generation until the proper fitness values achieved. Parents in the original algorithm are composed of the n number of individuals selected by a selecting algorithm, while parents in queen-bee evolution consist of the n/2 number of couples of a queen-bee. All parts of the individuals in LGP are mutated into small mutation probability while only some parts of individuals in LGP_BA are mutated into normal probability and others with strong probability.

The first feature of queen-bee evaluation fortify the extraction of LGP which is offspring and it is strongly depends on individual fittest and crossover operations, therefore, it also can increase the probability of the convergence. The next feature increase the exploration of LGPs with strong mutation and it can help to search new space of LGPs. These features makes LGPs to generates good solutions and evolve quickly. In the last step, queen-bee structure makes the possibility for LGPs to achieve an optimum response as decreasing the probability of convergence.

1.4. Support Vector Machine

This research has exploited SVM as the arranger in the wrapper specifications choice technique for evaluating feature subcategory. SVM known as a supervised machine learning method (Horng *et al*., 2011; Kahaki and Nordin, 2011a). The SVM classification is based on developing the hyper plane of n-dimensional data that create different categories of data. Based on the labeled data, SVM will be trained to recognize all possible solutions and the rest of data will be tested to determine which possibility is suitable to them at most. The training sets of data samples contains pairs of {(x,y)}, where x dependent on y. by maximize the SVM margin, the value of the best performance can achieve in terms of the classification process.

Equation 2 illustrating the minimize boundary SVM function.

Minimize:

$$W(\alpha) = -\sum_{i=1}^{l}\alpha_i + \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j k(x_i x_j) \quad (2)$$

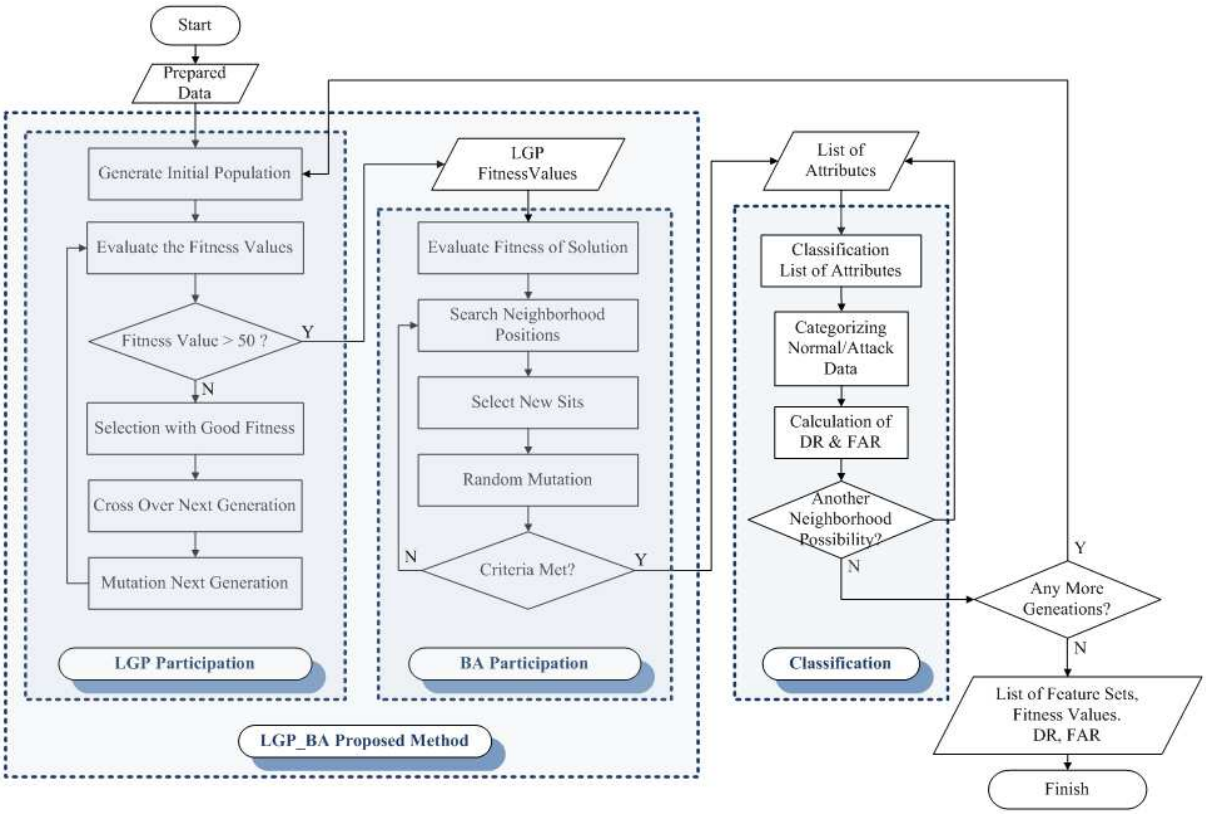

Fig 1. LGP_BA Flowchart (Enhancing LGP with BA)

Subject to:

$$\forall i : 0 \le \alpha_i \le C \text{and} \sum_{i=1}^{l} \alpha_i y_i = 0$$

where, l consider as the number which shows the count of training input data and α defined a vector variables of l. α can control the influence of the noise in the input data. The main SVM boundary function can define using support vectors which are the closest points to the class boundary. A vector α computed where each elements of α describes a weight of the input data points, where (α_i>0) and each α_i is the support values. The points with (α_l = 0) called none-support vectors while to determine the boundary function of SVM, support vector values can only be used. A data mapping to a higher dimensional space apply on all training data set in SVM which can be use to find a maximal margin of hyper plane separating values in SVM calculations.

1.5. Calculation of DR and FAR

To evaluate the proposed method, two evaluation metrics which are Detection Rate (DR) and False Alarm Rate (FAR) (Kahaki and Nordin, 2011b) has been selected which can be calculated using Equation (3 and 4). Recursive Operating Characteristic (ROC) is additionally calculated to prove the robustness of the method:

$$DR = \frac{\text{No. of intrusive correctly classified as intrusion}}{\text{Total No. of intrusion in the dataset}} \quad (3)$$

$$FAR = \frac{\text{No of normal correctly classified as intrusion}}{\text{Total No. of normals in the dataset}} \quad (4)$$

1.6. LGP_BA Performance

The experimental result of this technique will be evaluated and a comparison of the proposed method and the other feature selection applied method such as LGP, BA, Rough Set, Rough-DPSO, MARS and SVDF. Best feature sets and fitness's running over whole data are demonstrated in **Table 1**. For better observation on results they are listed from highest to lowest numerical values. From the **Table 1** it is possible to say that the features are involved in training process are C, L, W, X, AA and AI.

Table 1. The detection rate and false alarm rate results for all approaches

		Dataset2		Dataset3		Dataset4	
Technique		DR	FAR	DR	FAR	DR	FAR
SVDF	(B,D,E,W,X & AG)	53.80	1.05	76.10	0.00	62.64	0.12
MARS	(E,X,AA,AG,AH & AI)	62.90	25.20	89.40	4.95	85.50	5.50
Rough Set	(D,E,W,X,AI & AJ)	50.60	0.70	75.46	0.00	68.08	0.08
Rough DPSO	(B,D,X,AA,AH & AI)	61.40	22.00	90.02	4.39	86.19	3.90
LGP	(C,E,L,AA,AE & AI)	82.00	28.40	94.24	5.10	96.43	18.20
BA	(C,L,X,Y,AF & AK)	81.50	12.62	93.42	0.90	95.75	0.16
LGP_BA	(C, L, W, X, AA& AI)	82.60	17.40	94.20	4.80	96.70	12.20

```
01 Read from data set
// LGP PARTICIPATION
02 Initialize population
03 Randomly select m feature from n (select the features)
04 Calculate the fitness value using equation 4
05 If fitness value > 50 then
06   Store the gene and fitness value
07   Go to line 13
08 else
09   ApplyCross Over Operator to the Chromosome
10   ApplyMutation Operator to the Chromosome
11   Go to line 04
12 End if

// BA PARTICIPATION
13   Search neighbourhood possibilities
14   For each neighbourhood possibilities
15     Select new site
16     Apply random mutation to the chromosome (modify feature set)
17     Calculate fitness value
18     Store in the LGP_BA Attributes

//SVM CLASSIFICATION
19     Train the training dataset based on the extracted features
20     Compute SVM Cross Validation Solution using equation 1
21     Measure the dataset 2,3,4 based on the SVM solution
22     Calculate DR based on equation 2
23     Calculate the FAR based on equation 3
24     Store the taken values
25   End for

26 If new generation is available from LGP
27   Goto line 03
28 Else
29   Save all data (feature set, fitness value, DR and FAR)
30 End if
31 Finish
```

Fig. 2. LGP_BA Pseudo Code

2. RESULTS AND DISCUSSION

In order to evaluate the proposed method, four sample datasets extracted from KDD-cup99 have been used. Three evaluation metrics have been used which are Fitness value, Detection Rate (DR) and False Alarm Rate (FAR).

2.1. Comparison between LGP, BA and LGP_BA Results

In this section LPG and LGP_BA algorithm's performance are figuring out over the three datasets. As shown in **Table 1**, there has been a slight and almost insensible improvement about detection rate but regarding to the false alarm rate we have a significant reduction in every dataset. Although the combination of two techniques BA and LGP exceeds both critical time and complexity features but achieved advancement is referring as kind of tradeoff between accuracy and complexity and can be chosen owning to the importance of application.

As presented in **Table 1** and as it expected, the detection rates of three methods LGP, BA and LGP_BA are quite similar to each other. In some cases like the first and third datasets LGP_BA is becoming slightly improved and at the second dataset the result is almost the same. **Figure 3** illustrates to clarify this variation.

The good achievement of this work is highlighted in **Fig. 4**, which is the improvement of LGP in term of false alarm rate by BA modifications. In dataset one the false alarm rate decreased from 28% to about 17%. Getting this result can be considered as a good achievement. Moreover, BA itself still wins the challenge with lowest FAR.

2.2. LGP_BA Contrast to other Approaches

According to (Alomari and Othman, 2012), the various data mining methods such as SVDF, MARS, Rough set, Rough-DPSO, LGP and BA are applied for the intrusion detection systems by using same dataset that been used in this work. The experimental result, which is illustrated in **Table 1**, shows that although, LGP had the highest detection rate in the other side LGP carries the highest false alarm rate as well. This because of having high possibilities of mistakes it is not considered as an effective algorithm. But this is noteworthy that the improved method, which is LGP_BA is able to decrease the fault possibilities in all datasets.

According to the result of other methods applied on Dataset 3 clearly, it can be found that this dataset returns less possibilities of having faults. Furthermore, the experimental result in dataset 3 shows the highest Detection rate and less false alarm rate. Like pervious datasets, the last dataset which is the fourth one, returns the expected result that is shown in **Fig. 5.**

LGP is robustness technique with regard to detection rate but recent result proves that LGP_BA can provide the similar results and often times exposes rather more powerful against LGP. Overly, with the evidence of graph illustrated and subject to detection rate case, it is possible to declare that LGP, LGP_BA and BA are more reliable and stronger rather than other proposed techniques.

The **Fig. 5** demonstrates all testing datasets respectively. LPG has always highest false alarm rate among all available techniques while MARS comes next and compare to our developed LGP_BA technique it takes third position however SVDF is the absolute and the most powerful in false alarm rate aspect. LGP_BA is not taking the best false alarm rate, but the improvement of LGP is quite thoughtfulness. In addition, BA still wins in term of high DR and low FAR, but the good point is LGP_BA wins in term of high detection rate and the in term of FAR became closer to BA.

2.3. Experimental Results of Dr Tolerances Based On Fitness Values

The process flow of IDS shows that the classifier uses the fitness function to classify the chromosome enclosed to the fitness taken from optimization method. Moreover, according to literature review, it is assumed by the past researches that high fitness function will get high accuracy for IDS. In other words, previous researchers declare that how much the fitness function is greater the detection rate will be higher as well.

Although, this fact is true, the experimental achievements of this work by passing all fitness values to classifier without their value consideration, shows that highness of DR in IDS in not necessarily depends on the fitness function. In fact, the experimental result of this work shows that in some cases the fitness functions might have a sudden increase or drop. Meanwhile, DR is not just a gradually increase based on the fitness value. For instance in the real world, although, getting higher marks in an exam directly depends on how much a student studied, there is no guaranty for the student who studied more, takes higher mark. Nevertheless, based on the **Fig. 6 and 7**, in general the graphs are linear but increases and drops are visibly illustrated. The figures show that highest fitness function will cause high detection rate.

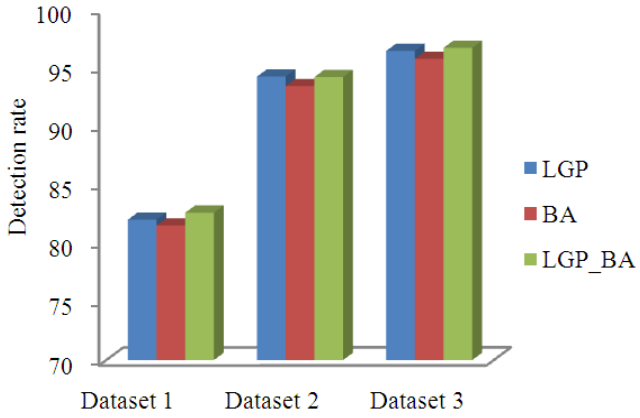

Fig. 3. Comparison between LGP, BA and LGP_BA method in detection rate

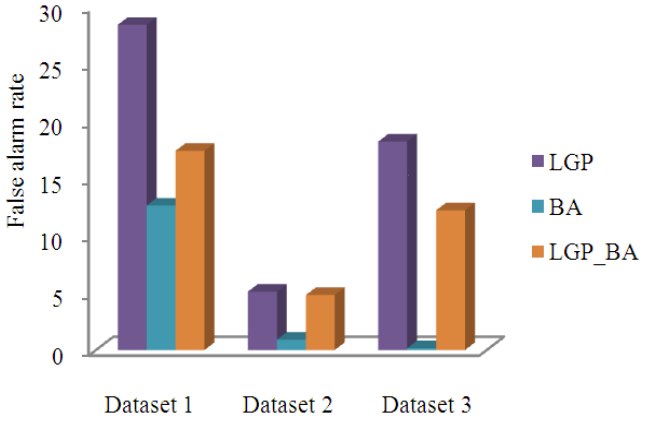

Fig. 4. Comparison between LGP, BA And LGP_BA Method In False Alarm Rate

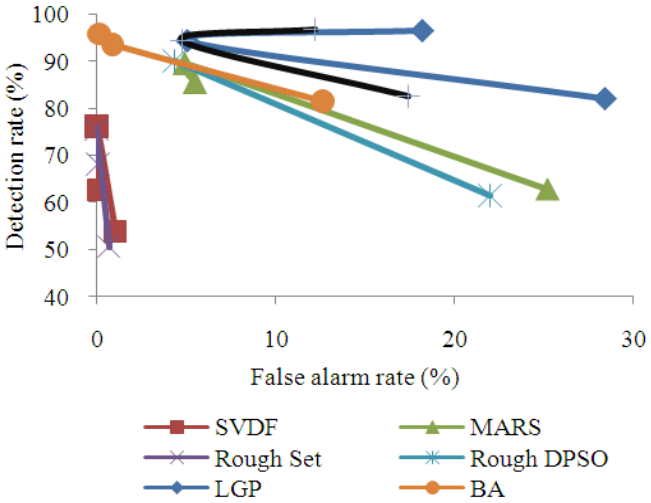

Fig. 5. The ROC graph for all applied methods in IDS

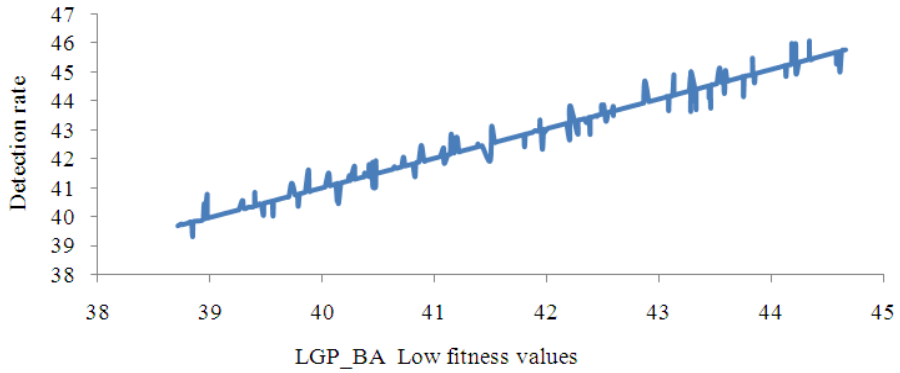

Fig. 6. Sample of DR tolerances dependencies of low fitness values

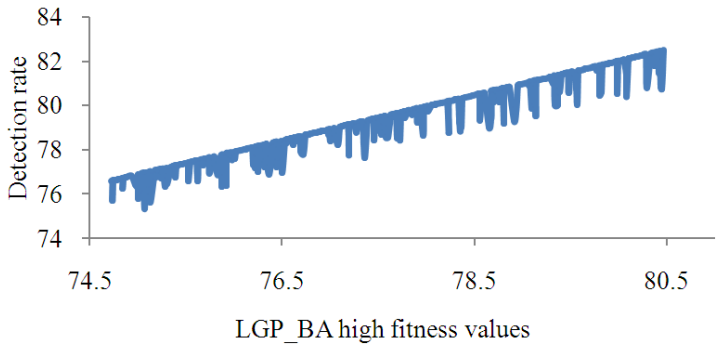

Fig. 7. Sample of DR tolerances dependencies to high fitness values

However, the dependencies between the ranges of fitness function are not equal. In addition, **Fig. 7** which contains the highest fitness values illustrates more drops rather than low fitness values that are shown in **Fig. 6.** It is noteworthy to mention that each of the following figures containing about 500 genes. Indeed, it is essential that if numbers of records or genes become higher the graph will be more linear.

The experiment result also shows that the DR of a feature set to the fitness value in the range of 40s, 50s or 60s is absolutely less than DR of a feature set with a fitness value of 70s or 80s. For instance, the DR of fitness value of 75 is always higher than the fitness value of 50. Moreover, the highest fitness value does not take the highest DR all the time. Furthermore, as can be seen in **Fig. 6 and 7** the stability of higher fitness values is less than lower ones. Besides, how the detection rate is higher; the ratio of increases in the DR is quite less rather than drops in DR. **Figure 7** shows this concept.

2.4. Statistical Test

For the statistical evaluation, independent sample Welch test is chosen. The value of alpha has been set to 0.05. As the result of t-test if value is less than alpha, it can be considered as significant. According to value of significant in the t-test result which are so close to zero; the result of t-test shows that, the significance level of LGP_BA result is high. Following subsections are showing the detail information of the t-test result over the LGP_BA result in comparison with LGP and BA method.

An independent-samples Welch-test was conducted to compare False Alarm Rate (FAR) in LGP, BA and LGP_BA methods. There was a significant difference in the scores for LGP (M = 21.87, SD = 3.31), BA (M=4.05, SD = 0.80) and LGP_BA (M = 4.96.01, SD = 1.01) conditions; the experimental result suggests that, FAR is significantly improved by LGP_BA in comparison with LGP but BA still wins in case of having lower FAR.As the output of the SPSS application of the significant test in term of DR, **Table 2** is shown grouped statistic and multiple comparison test of FAR for the three methods.

An independent-samples Welch-test was conducted to compare False Alarm Rate (FAR) in LGP, BA and LGP_BA methods. There was a significant difference in the scores for LGP (M = 21.87, SD = 3.31), BA (M = 4.05, SD = 0.80) and LGP_BA (M = 4.96.01, SD = 1.01) conditions; the experimental result suggests that, FAR is

significantly improved by LGP_BA in comparison with LGP but BA still wins in case of having lower FAR.

Table 2. Multiple Comparisons for Detection Rate (DR) among LGP, BA and LGP_BA methods by Welch-test

					95% Confidence interval	
(I) method	(J) method	Mean difference (I-J)	Std. Error	Sig.	*Lower bound*	*Upper bound*
LGP	BA	4.28161819	0.26869043	0.000	3.6398990	4.9233373
	LGP_BA	2.54768279	0.28009383	0.000	1.8787232	3.2166424
BA	LGP	-4.28161819	0.26869043	0.000	-4.9233373	-3.6398990
	LGP_BA	-1.73393540	0.29207272	0.000	-2.4315002	-1.0363706
LGP_BA	LGP	-2.54768279	0.28009383	0.000	-3.2166424	-1.8787232
	BA	1.73393540	0.29207272	0.000	1.0363706	2.4315002

Table 3. Multiple Comparisons for False Alarm Rate (FAR) among LGP, BA and LGP_BA methods by welch-test

					95% Confidence interval	
(I) method	(J) method	Mean difference (I-J)	Std. Error	Sig.	*Lower bound*	*Upper bound*
LGP	BA	17.82717416	0.05806137	0.000	17.6884785	17.9658698
	LGP_BA	16.91241051	0.05908349	0.000	16.7712770	17.0535440
BA	LGP	-17.82717416	0.05806137	0.000	-17.9658698	-17.6884785
	LGP_BA	-0.91476365	0.02215829	0.000	-0.9676856	-0.8618417
LGP_BA	LGP	-16.91241051	0.05908349	0.000	-17.0535440	-16.7712770
	BA	0.91476365	0.02215829	0.000	0.8618417	0.9676856

As the output of the SPSS application of the significant test in term of DR, **Table 3** is shown grouped statistic and multiple comparison test of FAR for the three methods.

3. CONCLUSION

The experiment result of this work shows that the proposed method is efficient to reduce the false alarm rate, besides the detection rate of proposed IDS becomes a bit grater. Furthermore, The SVM classifier shows that there is a not 100% dependency between fitness values and detection rate. Moreover, detection rate may have some increase and drop rather than the fitness ratio. In conclude the LGP_BA Optimization method is utilized to prove that GA and BA are working properly and there may have some expectation for other platforms. This research has proposed LGP_BA algorithm for IDS. The hypothesis of proposed algorithm has been justified the latest state of the art research in IDS. The results also have been evaluated using several methods based on DR, FAR and ROC. The result is compared with the BA and LGP and the other previous FS algorithms. According to the available references in the IDS area, so far many researchers proposed different approaches to detect attacks in the networks. In a nut shell, in the current studies all of the improvement by available techniques does not return much difference. Recently, there are more potential algorithms can produce good optimization techniques that can be apply in the feature selection. This work is based on combination of two strong algorithms and achieved acceptable results in this area. New algorithms such as Cuckoo Algorithm (Yang and Deb, 2009; Rajabioun 2011), water flow algorithm, Bat algorithm (Yang, 2010; 2011) are good for optimization that can be applied in IDS.

4. REFERENCES

Abraham, A. and D. Steinberg, 2001. MARS: Still an Alien Planet in Soft Computing? Computational Science-ICCS 2001, Alexandrov, V. (Ed.)., Springer, Berlin, ISBN-10: 3540422323, pp: 235-244.

Abraham, A., 2001. Neuro Fuzzy Systems: Sate-of-The-art Modeling Techniques. In: Connectionist Models of Neurons, Learning Processes and Artificial Intelligence, Mira, J. and A. Prieto (Eds.)., Springer-Verlag, Berlin, ISBN-10: 3540422358, pp: 269-276.

Alomari, O. and Z. Othman, 2012. Bees algorithm for feature selection in network anomaly detection. J. Applied Sci. Res., 8:1748-1756.

Anderson, L., T.F. Lunt, T. Javitz and H.S. Tamaru, 1995. Detecting unusual program behavior using the statistical components of NIDES. SRI International 333 Ravenswood Avenue, Menlo Park, CA.

Banzhaf, W., P. Nordin, R.E. Keller and F.D. Francone, 1998. Genetic Programming: An Introduction on the Automatic Evolution of Computer Programs and Its Applications. 1st Edn., Morgan Kaufmann Publishers, San Francisco, ISBN-10: 155860510X, pp: 470.

Baranidharan, T. and D.K. Ghosh, 2012. Medical image classification using genetic optimized elman network. Am. J. Applied Sci., 9: 123-126. DOI: 10.3844/ajassp.2012.123.126

Chavan, S., K. Shah, N. Dave, S. Mukherjee and A. Abraham *et al*., 2004. Adaptive neuro-fuzzy intrusion detection systems. Proceedings of the IEEE International Conference on Information Technology: Coding and Computing, Apr. 5-7, IEEE Xplore Press, pp: 70-74. DOI: 10.1109/ITCC.2004.1286428

Chebrolu, S., A. Abraham and J.P. Thomas, 2004. Hybrid Feature Selection for Modeling Intrusion Detection Systems. In: Neural Information Processing, Pal, N.R., N. Kasabov, R.K. Mudi, S. Pal and S.K. Parui (Eds.)., Springer, Berlin, ISBN-10: 3540239316, pp: 1020-1025.

Chen, Y., Y. Li, X.Q. Cheng and L. Guo, 2006. Survey and Taxonomy of Feature Selection Algorithms in Intrusion Detection System. Proceedings of the 2nd SKLOIS conference on Information Security and Cryptology, Nov. 29-Dec. 1, Springer Berlin Heidelberg, Beijing, China, pp: 153-167.

Cho, S.B. and H.J. Park, 2003. Efficient anomaly detection by modeling privilege flows using hidden Markov model. Comput. Security, 22:45-55. DOI:10.1016/S0167-4048(03)00112-3

Debar, H., M. Becker and D. Siboni, 1992. A neural network component for an intrusion detection system. proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy, May 4-6, IEEE Xplore Press, Oakland, CA., pp: 240-250. DOI: 10.1109/RISP.1992.213257

Folorunso, O., O.O. Akande, A.O. Ogunde and O.R. Vincent 2010. ID-SOMGA: A self organising migrating genetic algorithm-based solution for intrusion detection. Comput. Inform. Sci., 3: 80-92.

Guo, Y., B. Wang, X. Zhao, X. Xie and L. Lin *et al*., 2010. Feature selection based on Rough set and modified genetic algorithm for intrusion detection. Proceedings of the IEEE 5th International Conference on Computer Science and Education, Aug. 24-27, IEEE Xplore Press, Hefei, pp: 1441-1446. DOI: 10.1109/ICCSE.2010.5593765

Hettich, SB and S.D. Bay, 1999. The UCI KDD Archive. University of California, Department of Information and Computer Science, Irvine, CA.

Horng, S.J., M.Y. Su,Y.H. Chen, T.W. Kao andR.J. Chen *et al*., 2011. A novel intrusion detection system based on hierarchical clustering and support vector machines. Expert Syst. Applic., 38:306-313. DOI: 10.1016/j.eswa.2010.06.066

Jaganathan, Y. and I. Vennila, 2013. An integrated framework based on texture features, cuckoo search and relevance vector machine for medical image retrieval system. Am. J. Applied Sci., 10: 1398-1412. DOI: 10.3844/ajassp.2013.1398.1412

Kahaki, S.M.M. and M.J. Nordin, 2011a. Highway traffic incident detection using high-resolution aerial remote sensing imagery. J. Comput. Sci., 7: 949-953. DOI: 10.3844/jcssp.2011.949.953

Kahaki, S.M.M. and M.J. Nordin, 2011b. Vision-based automatic incident detection system using image sequences for intersections. Proceedigs of the International Conference on Pattern Analysis and Intelligent Robotics, Jun. 28-29, IEEE Xplore Press, Putrajaya, pp: 3-7. DOI: 10.1109/ICPAIR.2011.5976902

Lazarevic, A., L. Ertoz, V. Kumar, A. Ozgur and J. Srivastava, 2003. A comparative study of anomaly detection schemes in network intrusion detection. Proceedings of the 3rd SIAM International Conference on Data Mining, (CDM' 03), SIAM.

Lee, W. and S.J. Stolfo, 2000. Data mining approaches for intrusion detection. Proceedings of the 7th Conference on USENIX Security Symposium, Jan. 26-29, San Antonio, Texas.

Lippmann, R.P. and R.K. Cunningham, 2000. Improving intrusion detection performance using keyword selection and neural networks. Comput. Netw., 34:597-603. DOI:10.1016/S1389-1286(00)00140-7

López-Pujalte, C., V.P. Guerrero-Bote and F.D. Moya-Anegón, 2003. Order-based fitness functions for genetic algorithms applied to relevance feedback. J. Am. Society Inform. Sci. Technol., 54:152-160. DOI: 10.1002/asi.10179

Lunt, T.F., A. Tamaru, F. Gilham, R. Jagannathan and C. Jalali *et al*., 1992. A real-time Intrusion-Detection Expert System (IDES). SRI International, Computer Science Laboratory.

Mukkamala, S., A.H. Sung and A. Abraham, 2005. Intrusion detection using an ensemble of intelligent paradigms. J. Netw. Comput. Appli., 28:167-182. DOI:10.1016/j.jnca.2004.01.003

Mukkamala, S., A.H. Sung, A. Abraham and V. Ramos, 2006. Intrusion Detection Systems Using Adaptive Regression Spines. In: Enterprise Information Systems VI, Seruca, I., J. Cordeiro, S. Hammoudi and J. Filipe (Eds.)., Springer, Dordrecht, ISBN-10: 1402036752, pp: 211-218.

Nirmala, B.J. and S. Gowri, 2013. Ameliorate fuzzy c-means: An ameliorate fuzzy c-means clustering algorithm for ct-lung image segmentation. Am. J. Applied Sci., 10: 1439-1447. DOI: 10.3844/ajassp.2013.1439.1447

Premanode, B., J. Vongprasert, N. Sopipan and C. Toumazou, 2013. A novel multiclass support vector machine algorithm using mean reversion and coefficient of variance. J. Math. Stat., 9: 208-218. DOI: 10.3844/jmssp.2013.208.218

Rahman, S.A., A.A. Bakar and Z.A.M. Hussein, 2009. Filter-wrapper approach to feature selection using RST-DPSO for mining protein function. Proceedings of the IEEE 2nd Conference on Data Mining and Optimization, Oct. 27-28, IEEE Xplore Press, Kajand, pp: 71-78. DOI: 10.1109/DMO.2009.5341906

Sequeira, K. and M. Zaki, 2002. ADMIT: Anomaly-based data mining for intrusions. Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Jul. 23-25, ACM New York, NY, USA., pp: 386-395. DOI: 10.1145/775047.775103

Trair, D., W. Ma, D. Sharma and T. Nguyen, 2007. Fuzzy vector quantization for network intrusion detection. Proceedings of the IEEE International Conference on Granular Computing, Nov. 2-4, IEEE Xplore Press, Fremont, CA., pp: 566-566. DOI: 10.1109/GrC.2007.124

Yang, X.S., 2010. A New Metaheuristic Bat-Inspired Algorithm. In: Nature Inspired Cooperative Strategies for Optimization, González, J.R., D.A. Pelta, C. Cruz, G. Terrazas and N. Krasnogor (Eds.)., Springer, New York, ISBN-10: 3642240933,pp: 65-74.

Yang, X.S., 2011. Bat algorithm for multi-objective optimisation. Int. J. Bio-Inspired Comput., 3: 267-274. DOI: 10.1504/IJBIC.2011.042259

Yang, X.S. and S. Deb, 2009. Cuckoo search via lévy flights. Proceedings of the World Congress on Nature amd Biologically Inspired Computing, Dec. 9-11, IEEE Xplore Press, Coimbatore, pp: 210-214. DOI: 10.1109/NABIC.2009.5393690