AENSI Journals

Journal of Applied Science and Agriculture

Journal home page: www.aensiweb.com/jasa/index.html

Great Deluge Algorithm Feature Selection for Network Intrusion Detection

Zulaiha Ali Othman, Lew Mei Theng, Suhaila Zainudin, Hafiz Mohd Sarim

Data Mining & Optimization Group (DMO), Centre of Artificial intelligence Technology (CAIT), Faculty of Computer Science, Universiti Kebangsaan Malaysia, Selangor, 43600 Malaysia

ARTICLE INFO

Article history:
Received 6 September 2013
Received in revised form 14 October 2013
Accepted 15 October 2013
Available online 23 November 2013

Key words:
IDS, Great Deluge algorithm, Feature selection, Anomaly detection.

ABSTRACT

Intrusion detection systems (IDSs) deal with large amounts of data containing irrelevant and/or redundant features. These features result in a slow training and testing process, heavy computational resources, and low detection accuracy. Features selection, therefore, is an important issue in IDSs. A reduced features set improves the system accuracy and speeds up the training and testing process considerably. In this paper propose a wrapper–based feature selection techniques by using Great Deluge algorithm (GDA) asthe search strategy to specify a candidate subset for evaluation, as well as using Support Vector Machine (SVM) as the classifier technique. The experiments used four random datasets collected from KDD-cup99. Each data set contains around 4000 records. The performance of the proposed technique has been evaluated based on classification accuracy by comparing with other feature selection techniques such as Bees Algorithm (BA), Rough-DPSO, Rough, Linear Genetic Programming (LGP), Support Vector Decision Function Ranking (SVDF), and Multivariate Regression Splines (MARS). The result shows that the feature subset produced by GDA yield high classification accuracy when compared with other techniques.

© 2013 AENSI Publisher All rights reserved.

Introduction

Intrusion detection systems are security management systems that are used to discover inappropriate, incorrect, or anomalous activities within computers or networks. With the rapid growth of the Internet, these malicious behaviors are increasing at a fast pace and at the same time the vulnerability of computer security is adding pressure to the user in many aspects. Hence, the development of intrusion detection system has been given the highest priority by government, research institutes and commercial corporations. The security tools such as anti-virus, firewall and intrusion detection systems (IDS) are some of the reliable solutions available to protect our computers from cyber criminals; of which IDS has set a perfect platform to defend the confidentiality, integrity and other security aspects of the cyber world (Folorunso, O., 2010).

Nowadays, different kinds of IDSs are introduced. Each having different features for finding the different kind of network attacks. Network Intrusion Detection Systems are placed at a strategic point or points within the network to monitor traffic to and from all devices on the network. Ideally one would scan all inbound and outbound traffics. The network intrusion detection systems are classified into two types. The first type is signature based intrusion detection and the second is anomaly behavior detecting based intrusion detection. A signature based IDS will monitor packets on the network and compare them against a database of signatures or patterns of known malicious threats (Saravanan, C., 2012). This is similar to the way most antivirus software detects malware. Whereas the anomaly behavior detecting based intrusion detection system produces a normal traffic profile and uses it to spot any anomalous traffic patterns and intrusion attempts (Dat, T., 2007). According to Lazarevic *et al*. (2003) anomaly detection is a key element of intrusion detection in which the perturbations of normal behavior will suggest the presence of intentionally or unintentionally induced attacks, faults, defects, etc.

The main purpose of IDS is to address the classification problems by detecting intrusions in normal audit data. Among the large amount of features, there may be some irrelevant features with poor prediction ability to the target patterns, and some of the features may be redundant due to the highly inter-correlation of the features (Guyon and A. Elisseeff, 2003). Therefore, how to select a meaningful subset of features from the network traffic data stream becomes a very important topic in producing a high quality IDS.

The feature selection techniques are mainly divided into two categories, filter and wrapper, as defined in the work of John *et al*. (1994). Based on the wrapper, filter and hybrid methods, a number of techniques have been applied to identify intrusion features. The wrapper model uses the predictive accuracy of a classifier as a means to evaluate the "goodness" of a feature set, while the filter model uses a measure such as information,

Corresponding Author: Zulaiha Ali Othman, Data Mining & Optimization Group (DMO), Centre of Artificial intelligence Technology (CAIT), Faculty of Computer Science, Universiti Kebangsaan Malaysia, Selangor, 43600 Malaysia

consistency, or distance measures to compute the relevance of a set of features (Xia, T., 2005). Wrapper method utilized machines learning algorithm to assess the reliability of features. The filter method, however, will never use the machine learning algorithm for filtering the inappropriate and redundant features, instead will use the primary characteristics of the training data to assess the significance of the features or feature by distance measure, correlation measures, consistency measures and information measure (Gao, H.H., 2005). The feature selection is aimed at enhancing the performance in terms of learning speed, accuracy of prediction and simplification of rules. It also focuses on visualizing the data for model selection, reduction of dimensionality and noise removal (Jun Wang, 2009).

In this paper, the Great Deluge algorithm (GDA) is used to implement feature selection and Support Vector Machine is used for Classification. This paper is organized as follows. Section 2 describes the literature studies for feature selection in intrusion detection. Section 3 discusses the classification technique used in this research and Section 4 discusses the feature selection technique applied in this paper, followed by a detailed discussion of feature selection technique. Section 6 describes the intrusion data used and Section 7 discusses the experimental results. Finally, we conclude our work and discuss the future work in the last section.

Related Work:

Several literatures have tried to solve the problems by figuring out important intrusion features through feature selection algorithms such as genetic algorithm by Xia *et al*. 2005, ant colony optimization by Gao *et al*. 2005, particle swarm intelligence by Wang *et al*. 2009, bees algorithm by Osama *et al*. 2011 and these algorithms have yielded positive results. Feature selection is one of the most important and frequently used techniques in data preprocessing for IDS. It reduces the number of features, removes irrelevant, redundant, or noisy data, and brings the immediate effects for IDS.

Zainal *et al*. (2006) have projected a rough set theory for discovering the imperative characters in constructing IDS, and moreover the rough set has also been used to categorize data. The test results reveal that the feature subsets proposed by the Rough set are more vigorous and perform reliably during the experiments.

Sung and Mukkamala (2005) ranked six significant features. They used three techniques and compared the performance of these techniques in terms of classification accuracy on the test data. Those techniques were Support Vector Decision Function Ranking (SVDF), Linear Genetic Programming (LGP) and Multivariate Regression Splines (MARS). For detail results, please refer to (Dong Seong Kim, 2003).

Zainal *et al*. (2007) have proposed the Rough Set and discrete Particle Swarm Optimisation (Rough-DPSO) for feature assortment in intrusion detection. The Rough-DPSO is a two-stage method that isolates the training dataset before it is fed into a Rough Set tool called Rosetta. GA was used for reducts. The most appeared 15 features are selected depending on the reducts generated and are treated as initial feature subset. This particular stage is considered as a very crucial process by researchers as it eliminates the insignificant and repeated features. The initial feature subset discussed earlier would later turn out to be an input to the subsequent stage (granular feature selection using Rough-DPSO). Particle swarm will search the space of feature subsets that are exponential with the number of features. Rough-DPSO has to examine only 15 features subset generated from Rough Set instead of all the 41 existing features.

Support Vector Machine (SVM):

Support Vector Machine (SVM) is based on the structural risk minimization principle from the statistical learning theory. Its kernel is to control the empirical risk and classification capacity in order to maximize the margin between the classes and minimize the true costs (Chen, Y., 2006). A support vector machine finds an optimal separating hyper-plane between members and non-members of a given class in a high dimension feature space (Liu, H., H. Motoda, 1998). Although the dimension of feature space is very large, it still shows good generalization performances. Pass research shown that SVM has shown as a best classification algorithm for IDS [Osama,…].

Great Deluge Algorithm:

The Great Deluge algorithm (GDA) was proposed by Dueck in 1993. It is similar in many ways to its predecessors (eg. Hill-climbing and Simulated Annealing (SA)). The main difference with the SA algorithm is the deterministic acceptance function of the neighboring solution. The inspiration of the GDA algorithm comes from the analogy that in a Great Deluge a person climbing a hill tries to avoid his feet wet in order to move in any direction in the way of finding a way up as the water level rises. Finding the global optimum of an optimization problem could be seen as finding the highest point in a landscape. The GDA accepts all solutions, for which the absolute values of the cost function are less than or equal to the current boundary value, called "level". The initial value of the level is equal to the initial objective function. The advantage of GDA is that it only depends on one parameter, called "up" value that represents the speed of the rain. The quality of the results is highly dependent on the value of the "up" parameter. If the value of the "up' is high, the algorithm will be fast

and will produce poor quality results. In other words, if the "up" value is small the algorithm will relatively generate better results within a higher computational time.

In (Burke, E.K., 2003) the authors extended GDA by accepting all downhill moves. This practice was successfully applied to optimize the university examination timetabling problems and its performance was thoroughly investigated. Further multi-objective modification of this algorithm have been introduced and evaluated in (Bykov, Y., 2003) and (Petrovic, S., Y. Bykov, 2003). All these investigations have yielded positive outcomes. The proper utilization of these properties significantly increases the performance of a local search. When compared experimentally with other existing techniques, GDA produced higher results on most benchmark problem instances (Burke, E.K., 2007). The following shows the pseudo-code of GDA as stated in (Mafarja, M. and S. Abdullah, 2011).

Great Deluge Algorithm

Generate a random initial solution *Sol;*
SetSol_{best}= *Sol;*
Set *level* = *f*(Sol_{best})*;*
Set estimated quality of every solution, *EstimatedQuality = 100*;
Set number of iteration, *NumOfIte_GD = 250*;
Calculate **increase rate**, *β= EstimatedQuality/NumOfIte_GD*;
Set *Iteration*← 0;
While (*Iteration*<*NumOfIte_GD*)
Generate at random a new solution Sol_{trial}in the neighbour of *Sol*
Calculate *f*(Sol_{trial})
if(*f*(Sol_{trial})> *f*(Sol_{best}))
Sol ←Sol_{trial} *;* Sol_{best}←Sol_{trial}
f(*Sol*) = *f*(Sol_{trial})*;* *f*(Sol_{best}) = *f*(Sol_{trial})
elseif(*f*(Sol_{trial})> *level*)
Sol←Sol_{trial} *;* *f*(*Sol*)← *f*(Sol_{trial})
end if
if(*f*(Sol_{trial}) ≥ *level*)
level = *level* + *β*
end if
Iteration++
*end while*Return Best Solution Found

Feature Selection:

In this paper, we have proposed GDA as feature selection approach for network anomaly detection. This approach uses the wrapper approach as a random search technique for subset generation and uses SVM as the classifier technique. Later, a few tests have been conducted to classify the normal and attack data. Feature selection using GDA consists of many components: solution representation, parameter setting, neighborhood structures and fitness function.

Solution Representation:

This research requires a data set for use in the feature selection process, and this data set constitutes patterns with *Nt* features. The complete set of features is represented by a binary string of length N, where a bit in the string is set to '1' if it is to be kept, and set to '0' if it is to be discarded, and N is the original number of features (Wang, J., 2010).

Neighbourhood structures:

Various operators could be employed to create neighborhood solution, which include the monadic operators such as mutation, inversion, swap, and insertion (single or multiple) (Muhamad, D.T., 2007). A classical Random Flip Mutation operator (RFM) have been considered for the purpose of the neighbourhood search, in which the randomly chosen *Nf* bits are turned over to configure the present neighbourhood as example shown in (Osama,).

Fitness function:

In this paper, SVM is used as the classifier in the wrapper approach feature selection method, where the SVM classifier evaluates every feature subset, and gauges the quality of the symbolized feature subset. The fitness of every feature subset is measured bymeans of 10-fold Cross Validation (10FCV). 10FCV is used o generate the accuracy of classification by a SVM that is accustomed to this feature subset. In the 10FCV, the data set contains 10 subsets, which one subset is used for testing and the remaining for training purposes (Osama Alomari, 2012). Next the average accuracy rate across all 10 trials is computed. The complete fitness function is described in (Zainal, A., 2007) as the following:

$$\alpha * \gamma R(D) + \beta * \frac{|C| - |R|}{|C|}$$

where γR(D) is the average of accuracy rate obtained by conducting ten multiple cross-validation with SVM classifiers on the training dataset, with attribute set R selected from feature subset in the population to decision D.|R| is the '1' number of position or the length of selected feature subset. |C| is the total number of features. α and β are two parameters corresponding to the importance of classification quality and subset length, $\alpha \in [0,1]$ and $\beta = (1-\alpha)$, respectively. The classification quality is more important than subset length.

The proposed GDA-SVM Feature Selection Approach:

The proposed GDA-SVM consists of six steps, which are discussed as the followings:

Step 1: (Initialization) Randomly generates an initial solution. For the GDA algorithm, the complete set of features is represented by a binary string of length N, where a bit in the string is set to '1' if it is to be kept, and set to '0' if it is to bed is carded, and N is the original number of features.

Step 2: Measure the fitness of the initial solution, where the accuracy of the SVM classification and all the chosen features are utilized to calculate the fitness function. Eq. 4 appraises the fitness of the initial solution.

Step 3: The algorithm conducts search in the neighbourhood of the initial solution. The neighbourhood search uses a mutation operator. A new random solution is generated.

Step 4: Eq. 4 evaluates the fitness of the new solution. Accept the solution where the fitness is equal or more than the level. Update the best solution if the fitness of the new solution is higher than the current best solution and level with a fix increase rate.

Step 5: Repeat these steps until a stopping criterion is met. If stopping condition is satisfied, solution with best fitness is chosen; otherwise, the algorithm will generate new solution. In this paper, number of iteration is assigned as the stopping criteria.

Step 6: At the end, train SVM based on the best feature subset. After that, testingis to be performed by using testing datasets.

The experiments have been carried out for ten times to validate the performance of the GDA feature selection technique. The pseudo-code of GDA is shown in Figure 3.

1. Initialize a random solution
2. Evaluate the fitness of the solution.
3. While (stopping criterion not met)
4. Generate at random a new solution in the neighborhood of initial solution.
5. Evaluate the fitness of the new solution. Accept the solution where the fitness is equal or more than level
6. End while
7. Output: best feature subset.

Fig. 3: Pseudo code of GDA-SVM approach for feature selection.

Intrusion Data:

This paper uses KDD-CUP 99 data subset that was pre-processed by the Columbia University and distributed as part of the UCI KDD Archive. The training dataset contains about 5 millions connection records and 10% of the training dataset consists of 494,021 connection records. In turn each record contains 41 features and a label which indicates the attack name. The label is only supported for the training dataset. The objective of the competition is to build a network intrusion detector, a predictive model or a classifier that can detect "bad" connections, called as intrusion or attacks, and identify "good" connections, called as normal connections (Shirazi, H.M., 2010). For easier referencing, each feature is assigned a label (A to AO). This referencing is adopted from (Chebrolu, S., 2005). Some of these features are derived features, where the features are either nominal or numeric.

In the training dataset there are 24 types of attacks. Each attack mayfall into the four main categories: 1) DOS: denial-of-service, e.g. syn flood. 2) R2L: unauthorized access from a remote machine, e.g. guessing password. 3) U2R: unauthorized access to local super user (root) privileges, e.g., various 'buffer overflow 'attacks. 4) Probing: surveillance and other probing, e.g., port scanning.

We have used the features used by Almori & Othman, 2011, and trained each of them using our training dataset. Our SVM classifier was trained based on each feature subsets. Then, we test using our testing datasets and their results were compared. Here, the detection could either be attack or normal.

Experimental and Results:

The GDA-SVM feature selection approach is implemented in Java. The experiment used a Core i3 CPU 2.10 GHZ with 4 GB RAM, Window 7 Home Edition. The SVM classifier used in this approach is based on the LIBSVM. The parameters for the SVM are C = 1, Gamma = 0.01, and the Gaussian radial basis function (RBF) is used as the kernel. Our experiments were performed on KDD Cup 1999 intrusion detection dataset, in order to evaluate our proposed feature selection approach. Our experiments have two phases, namely, training and testing phases. This dataset is again divided into 4 sets of data. One dataset is used for training and the remaining is used for testing. Each set has 4000 randomly chosen records, where nearly half of the data (50 to 55%) belong to the normal category and the leftovers are mere attacks. Moreover, the experiments for each dataset are conducted for ten times to evaluate the performance of the proposed approach.

Parameters Settings:

The parameters used in the GDA algorithm is shown in Table 2. The values were decided empirically. Several feature selections have been applied in the intrusion detection. As mentioned earlier, (Dong Seong Kim, 2003) had suggested 3 feature subsets that were produced by three techniques. Zainal *et al*. had also suggested 2 feature subsets produced by two techniques, which are Rough Set (Zhang Kun, 2006) and Rough-DPSO (Pratik N., 2011).

By using features that are proposed in the above, the training has been conducted in order to train each of the features using the training dataset. The SVM classifier was trained based on each feature subsets. Then, three testing datasets were used to test the feature subsets and their results were compared.

Table 2: GDA Parameter Setting.

Parameter	Estimated Quality	Number of iteration	Increase Rate
Values	100	100	1.0

GDA Performance based on Highest Fitness Function:

The first experiment conducted follows Almori & Othman, as it was stated that the highest fitness function resulting in the highest accuracy. The best feature subset acquired from the second run, constituted the highest fitness and was used to train the SVM classifier (libsvm). The features involved in the training process are: are B, G, H, J, N, S, W, G and L. Later a test was conducted to classify the normal and attack data. The classification accuracy was used as the parameters for evaluating the efficiency of GDA.

Table-3 compares the classification performance for all the seven feature subsets produced by previous techniques. The first three rows are the feature subsets proposed by (Pratik N., 2011). The fourth and five rows are the features proposed by Rough Set and Rough-DPSO. The seventh row is proposed by (Osama Alomari, 2012). The last row are the nine significant features proposed by our approach. The last column shows the mean value of each technique. The mean gives the average performance of the feature subset proposed by the respective technique on three different test sets.

Table 3: Comparison of Classification Rate.

Technique	Data2	Data3	Data4	Mean
LGP(C,E,L,AA,AE&AI)	75.37	94.60	87.65	85.87
SVDF(B,D,E,W,X &AG)	82.60	89.83	84.98	85.80
MARS(E,X,AA,AG,AH&AI)	70.47	92.65	90.87	84.66
Rough Set(D, E, W, X, AI & AJ))	81.65	89.56	87.18	86.13
Rough-DPSO(B,D,X,AA, AH & AI)	71.85	93.23	92.07	85.72
BA(C, L,X,Y,AF&AK)	85.25	96.65	98.20	93.36
GDA(B,G,H,J,N,S,W,G,L)	83.16	90.75	87.45	87.12

As shown in Table 3, it can be observed that the feature subset produced by BA feature selection approach still yield the highest classification accuracy for all dataset. In dataset 2, GDA yields the second highest. BA performs 2.09% classification accuracy better compared to GDA. In dataset 3, LGP presents the second highest. BA yields better by 2.05% compared to LGP. However GDA is fifth in the ranking. In dataset 3, Rough-DPSO is ranked second. The result of BA is better compared to Rough-DPSO by 6.13% of the classification accuracy. Looking at the mean values for each technique, GDA has the second highest average classification rate. GDA has outperforms other techniques except BA. Furthermore, BA has the highest average classification rate.

GDA Performance not based on Highest Fitness Function:

The second experiment was continued using all fitness function. The experiment result shown in Table 4 shows that the highest fitness function does not necessary resulting in the highest classification accuracy The feature subset acquired from the second run acquired the highest fitness function among dataset 1 which is 71.56% and its mean classification accuracy is 87.13%. However, in the eighth run, it acquired the highest mean classification accuracy of90.38% and its fitness function is only 71.25%. The fitness function in the eighth run

is0.31% lower than the second run. Therefore, we can conclude that highest fitness function does not necessary represent the best classification accuracy.

In this experiment GDA yields an average of 90.30% accuracy compared to using the highest fitness function, which is87.12%. GDA also yields the best classification accuracy 95.68% in dataset 3, with the features of CF,K,L,P,Q,R,U,Z,AB,AC,AI,AJ,AM,AN and AO. However, BA still remains the best with accuracy of 98.20% in dataset 4.

Table 4: Comparison of classification accuracy versus fitness function.

No	Best Feature (Data set 1)	Fitness Value	Classification Accuracy			
			Data Set 2	Data Set 3	Data Set 4	Mean
1	(B,C,H,P,Z,AC)	68.76	80.30	92.46	91.88	88.21
2	(B,G,H,J,N,S,W,G,L)	71.56	83.18	90.76	87.46	87.13
3	(I,K,N,U,Y,AA,AE,AL)	59.39	72.05	79.97	71.84	74.62
4	(N,Q,S,U,AC,AE,AO)	61.58	70.83	77.82	75.56	74.74
5	(D,E,V,AC,AF,AN)	62.03	71.55	77.87	75.79	75.07
6	(H,L,R,V,AH)	63.34	75.38	94.56	87.63	85.85
7	(R,W,AB,AJ,AK,AO)	63.99	75.90	82.34	79.41	79.22
8	(K,V,X,Y,Z,AH,AJ)	65.76	73.08	88.76	88.18	83.34
9	(CF,K,L,P,Q,R,U,Z,AB,AC,AI,AJ,AM,AN,AO)	71.25	81.45	95.68	94.00	90.38
10	(J,P,X,AB,AK,AL,AN)	67.72	79.95	92.38	81.86	84.73

Performance GDA in Other Data Set:

The previous experiments were only conducted on dataset 1 as training and datasets 2, 3 and 4 as testing. In order to have robust experiments, data set 2, 3 and 4 also used as training whilst other datasetsare used as testing.16 experiments were conducted. Table 5 shows the result of the experiments obtained by GDA in these dataset.

Table 5: Classification Accuracy using the Best GDA Features Selection

Training Data Set	Testing Data Set	GDA (%)	
		The best Accuracy	MeanAccuracy
Data set 1 (CF,K,L,P,Q,R,U,Z,AB,AC,AI,AJ,AM,AN,AO	Data set 2	81.45	
	Data set 3	95.68	90.38
	Data set 4	94.00	
Data set 2 (B,F,J,P,R,Y,AF,AG,AI,AK,AL,AM,AO)	Data set 1	80.70	
	Data set 3	95.20	91.24
	Data set 4	97.83[b]	
Data set 3 (R,S,W,AA,AG,AL,AN)	Data set 1	77.13	
	Data set 2	81.65	81.26
	Data set 4	85.01	
Data set 4 (O,S,W,X,AC,AF,AH,AI)	Data set 1	77.78	
	Data set 2	80.98	83.04
	Data set 3	90.38	

The result shows that dataset 2 obtained the best accuracy of 97.83%, with means of 91.24% compared to the previous research, which only focused on dataset 1 as training (95.68% with means 90.38%).

Figure 4 shows the performance of the features selection algorithm performed by IDS based on accuracy (in %). The figure shows that applying GDA does not result in better performance than Bees Algorithm. However, the accuracy of GDA has increased from 89.12% to 90.35%by conducting the experiment using all the fitness function and not selecting the highest fitness function. The figure also shows that by applying robust experiment on the datasets as training and testing has increased the average accuracy up to 91.24%.

Performance of GDA using ROC:

The ROC curve is a method to visualise the trade-offs between DR and FAR in intrusion detection model. ROC is a method to select the best result among the features selection in a dataset. The ROC graph represents y= f(x) in (x,y) coordinates, where according to Provost and Fawcett [23],the upper left point (0,1) represents the ideal IDS, which has a 100% detection rate and 0% false-alarm rate, respectively.

DR and FAR are calculated based on the accuracy result from the equation stated in Equation 4. Table 6 presents the comparison results between the feature selection techniques GDA, BA, BSA, LGP, SVDF, MARS, Rough Set, Markov_Blanket and Rough_DPSO. As mentioned earlier, dataset 2, 3 and 4 are used to evaluate the proposed techniques and other existing feature selection techniques. Figure 5 shows the ROC graph of BA,LGP,SVDF,MARS, Rough Set, Markov_Blanket and Rough_DPSO techniques in dataset 2, 3 and 4. From the graph it can be seen that GDA is the nearest to the upper left corner of ROC. It has achieved the average of 90.84% detection rate and 0.17% false alarm rate. Based on ROC graph,GDA has outperformed all other techniques.

Journal of Applied Science and Agriculture, 8(4) Sep 2013, Pages: 322-330

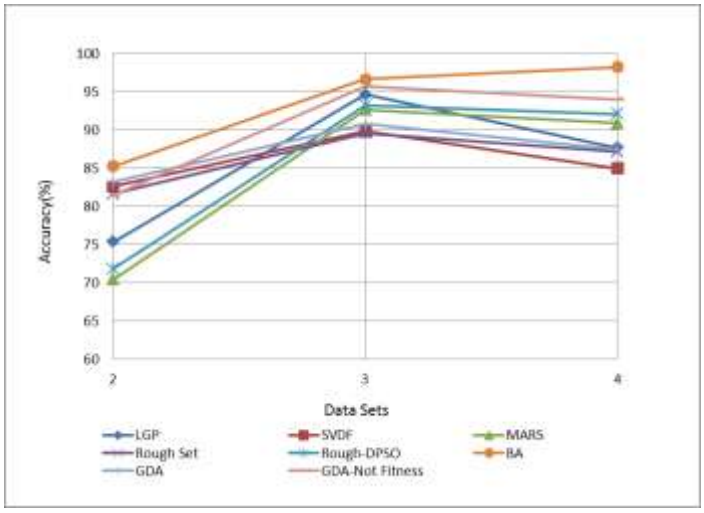

Fig. 4: Features Selection Algorithm Performance based on Accuracy(%).

Table 6: DR and FAR Results.

	Data Set 2		Data Set 3		Data Set 4		Average	
Technique	DR	FAR	DR	FAR	DR	FAR	DR	FAR
LGP	82	28.4	94.24	5.1	96.43	18.2	90.89	17.23
SVDF	53.8	1.05	76.1	0	62.64	0.12	64.18	0.39
MARS	62.9	25.2	89.4	4.95	85.5	5.5	79.27	11.88
Rough Set	50.6	0.7	75.46	0	68.08	0.08	64.71	0.26
Markov Model	56.41	1.41	79.6	0	68.9	0.04	68.3	0.48
Rough-DPSO	61.4	22	90.02	4.39	86.19	3.9	79.2	10.1
BA	81.5	12.62	93.42	0.9	95.75	0.16	90.22	4.56
BSA	76.41	4.03	87.8	0.26	83.26	0.08	82.49	1.46
GDA	55.03	0.01	78.29	0.00	68.64	0.00	55.03	0.01
GDA Not Fitness	82	0.28	94.13	0.05	96.38	0.18	90.84	0.17

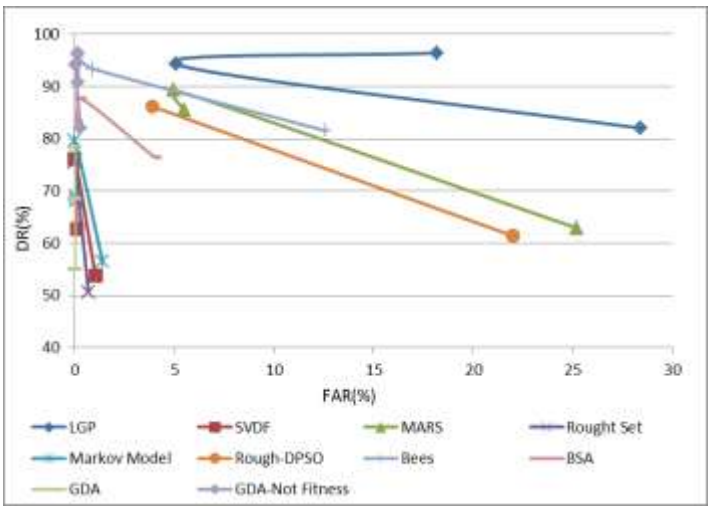

Fig. 5: ROC Graph for all techniques.

Although BA has achieved better results compared with GDA in term of classification accuracy, the average detection rate for BA in ROC is 4.56%, which is 4.39% higher than GDA and its detection rate is 90.22%, which is 0.62% lower than GDA. Based on the ROC graph, the results indicate that the feature subset proposed by GDA algorithm is superior in terms of detection rate, false alarm rate and robustness.

Table-7 shows the best performance trade-off between the training dataset and testing dataset, and Figure 6 shows the comparative performance ROC graph of GDA. The ROC graph are plotted to dissipate the relationship between the false alarm rate and the detection rate, in order to show the trade-off between them. The best performance for each training and testing results are used to plot the ROC graph. Out of 12 sets of

results, the performance for training dataset 2 against testing dataset 4 during sixth run has outperformed others. It has achieved 96.38% in detection rate and 0.16% in false alarm rate.

Based on the datasets used for the experiments, the results indicate that the feature subset generated by dataset 2 is superior in term of the detection rate, false alarm rate and robustness.

Table 7: DR and FAR for GDA versus BA.

Training Data Set	Testing Data Set	GDA (%)	
		DR	FAR
Data set 1	Data set 2	82.00	0.28
	Data set 3	94.13	0.05
	Data set 4	96.38	0.18
Data set 2	Data set 1	77.65	0.34
	Data set 3	94.25	0.05
	Data set 4	96.38	0.16
Data set 3	Data set 1	72.06	0.25
	Data set 2	74.62	0.27
	Data set 4	85.51	0.43
Data set 4	Data set 1	75.59	0.24
	Data set 2	80.76	0.24
	Data set 3	89.44	0.31

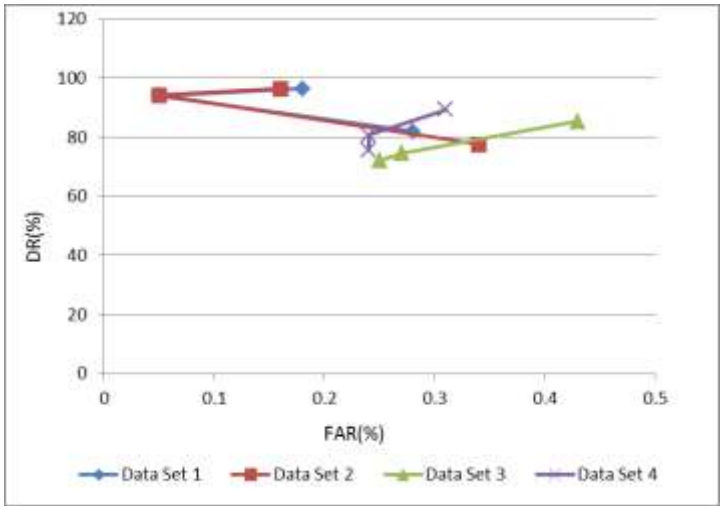

Fig. 5: ROC Graph for GDA.

Conclusion and Future Work:

In this paper, we have proposed GDA-SVM based wrapper approach for feature subset selection for anomaly detection. Based on the datasets used for the experiments, the results have proved that the feature subset proposed by GDA has not outperformed BA. Thus, it can be said that the feature subset produced by BA yields better quality IDS compared to GDA. We also can conclude here that the highest fitness function does not necessary always present the better classification accuracy. Besides that, out of the four datasets, the best intrusion detection model using ROC measurement was obtained only from dataset 2. This research concludes that GDA has been identified as one useful feature selection for IDS. Furthermore, to produce better accuracy for IDS, need to conduct robust experiments. Assumption of having highest fitness function does not always obtained highest accuracy.

References

Burke, E.K., Y. Bykov, J. Hirst, 2007. “Great Deluge Algorithm for Protein Structure Prediction”.

Burke, E.K., Y. Bykov, J.P. Newall, S. Petrovic, 2003. “A Time-Predefined Local Search Approach to Exam Timetabling Problems”. Accepted for publication in*IIE transactions on Operations Engineering*.

Bykov, Y., 2003. Time-Predefined and Trajectory-Based Search: Single andMultiobjective Approaches to Exam Timetabling. PhD Thesis. The Universityof Nottingham, Nottingham, UK.

Chebrolu, S., A. Abraham, J.P. Thomas, 2005. Feature Deduction and Ensemble Design of Intrusion Detection Systems. Journal of Computers and Security, 24(4): 295-307.

Chen, Y., Y. Li, X.Q. Cheng, L. Guo, 2006. Survey and taxonomy of feature selection algorithms in intrusion detection system, pp: 153-167.

Dat, T., M. Wanli, D. Sharma, N. Thien, 2007. Fuzzy Vector Quantization for Network Intrusion Detection. Granular Computing, GRC 2007. IEEE International Conference on,pp. 566-566.

Dong Seong Kim, Jong Sou Park, 2003. Network-based intrusion detection with support vector machines.In:Kahng H-K. Ed. ICOIN 2003, LNCS 2662: 747-756.

Dueck, G., 1993. "New Optimization Heuristics. The Great Deluge Algorithm and theRecord-to-Record Travel". *Journal of Computational Physics*, 104: 86-92.

Folorunso, O., O.O. Akande, A.O. Ogunde, O.R. Vincent, 2010. "ID-SOMGA: A Self Organising Migrating Genetic Algorithm-Based Solution for Intrusion Detection". Computer and Information Science, 3(4): 80-92.

Gao, H.H., H.H. Yang, X.Y. Wang, 2005. Ant colony optimization based network intrusion feature selection and detection, Machine Learning and Cybernetics, Proceedings of 2005 International Conference on, 6: 3871-3875.

Guyon and A. Elisseeff, 2003. "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, 3: 1157-1182.

http://kdd.ics.uci.edu/databases/kddcup1999/kddcup1999.html.

John, G., R. Kohavi and K. Pfleger, 1994. "Irrelevant Features and the Subset Selection Problem," in *Proceedings ML-94*: 121-129, Morgan Kaufmann.

Jun Wang, Xu Hong, Rong-rong Ren and Tai-hang Li, 2009. A Real-time Intrusion Detection System Based on PSO-SVM.International Workshop on Information Security and Application, pp: 319-321.

Lazarevic, A., L. Ertoz, V. Kumar, A. Ozgur, J. Srivastava, 2003. A comparative study of anomalydetection schemes in network intrusion detection, Proceedings of the third SIAM International Conference on Data Mining, 3.

Liu, H., H. Motoda, 1998. "Feature Selection for Knowledge Discovery and Data Mining", Boston: Kluwer Academic.

Mafarja, M. and S. Abdullah, 2011. Modified great deluge for attribute reduction in rough set theory.Fuzzy Systems and Knowledge Discovery, pp: 1464-1469.

Muhamad, D.T., Z.M. Mahmuddin, A. Ghanbarzadeh, E. Koc, S. Otri, 2007c. Using the bees algorithm to optimise a support vector machine for wood defect classification. *In:IPROMS 2007 innovative production machines and systems virtual conference*, Cardiff, UK.

Osama Alomari, Zulaiha Ali Othman, 2012. Bees Algorithm for feature selection in Network Anomaly Detection, Journal of Applied Sciences Research, 8(3): 1748-1756.

Petrovic, S., Y. Bykov, 2003. "A Multiobjective Optimisation Technique for ExamTimetabling Based on Trajectories". E. Burke, P. De Causmaecker (eds.), *ThePractice and Theory of Automated Timetabling IV: Selected Papers (PATAT2002).* Lecture Notes in Computer Science, Springer-Verlag, Berlin,Heidelberg, New York, 179-192.

Pratik N., Neelakantan, C. Nagesh M. Tech, 2011. "Role of Feature Selection in Intrusion Detection Systems for 802.00 Networks" *International Journal of Smart Sensors and Ad Hoc Networks(IJSSAN)* 1(1).

Saravanan, C., M.V. Shivsankar, P. Tamije Selvy, S. Anto, 2012. An Optimized Feature Selection for Intrusion Detection using Layered Condition Random Fields with MAFS. International Jornal of Mobile Network Communications & Telematics (IJMNCT), 2(3).

Shirazi, H.M., 2010. An Intelligent Intrusion Detection System Using Genetic Algorithms and Features Selection. *Majlesi Journal of Electrical Engineering*, 4(1).

Sung, A., S. Mukkamala, 2005. The feature selection and intrusion detection problems. Advances in Computer Science-ASIAN 2004.Proceedings of Third SIAM Conference on Data Mining, San Francisco, pp: 3192-3193.

Wang, J., T. Li, R. Ren, 2010. A real time IDSs based on artificial Bee Colony-support vector machine algorithm. Third International Workshop on Advanced Computational Intelligence, pp: 91-96.

Xia, T., G. Qu, S. Hariri, M. Yousif, 2005. An efficient network intrusion detection method based on information theory and genetic algorithm. Performance, Computing, and Communications Conference, 2005. IPCCC 2005. 24th IEEE International, pp: 11-17.

Zainal, A., M. Maarof, S. Shamsuddin, 2007. Feature selection using rough-DPSO in anomaly intrusion detection. Computational Science and Its Applications–ICCSA, pp: 512-524.

Zainal, A., M. Maarof, S. Shamsuddin, 2007. Feature selection using rough-DPSO in anomaly intrusion detection. Computational Science and Its Applications–ICCSA, pp: 512-524.

Zainal, A., M.A. Maarof, S.M. Shamsuddin, 2006. Feature Selection Using Rough Set in Intrusion Detection. TENCON 2006. 2006 IEEE Region 10 Conference, pp: 1-4.

Zhang Kun, Cao Hong-xin, Yan Han, 2006. Application of support vector machines on network abnormal intrusion detection. Application Research of computers, 5: 98-100.