Dynamic Ant Colony System with Three Level Update Feature Selection for Intrusion Detection

Helmi Md Rais and Tahir Mehmood
(Corresponding author: Tahir Mehmood)

Department of Computer and Information Science, Universiti Teknologi Petronas
32610 Seri Iskandar, Perak, Darul Ridzuan, Malaysia.
(Email: tahirmehmood.seecs@gmail.com)
(Received Oct. 4, 2016; revised and accepted Feb. 2 & Mar. 11, 2017)

Abstract

The current era is known as the age of digital information and general medium of access to this information is computer networks. The uses of network technology also make information insecure. Intrusion Detection System (IDS) has been proven effective against such attacks. The anomaly-based detection method is good to detect new attacks. One of the foremost shortcomings in the anomaly-based detection is the irrelevant and redundant features to the classification algorithm that results in low detection rate. Therefore, the primary objective of the feature selection process is to enhance the classification accuracy by removing redundant and irrelevant features. In this research a new feature selection algorithm called, Dynamic Ant Colony System with Three Level Update Feature Selection, has been proposed. The proposed method uses a different level of pheromones that help ants to find the robust features. The method also utilizes the information of each individual ant during feature selection process and incorporates the accuracy of the classification algorithms. Results showed that proposed feature selection algorithm outperformed compared to the previous feature selection algorithms.

Keywords: Ant Colony Optimization; Feature Selection; Intrusion Detection System

1 Introduction

Network security is becoming a crucial and elementary task for the organizations. Due to which many tools are being developed to overcome the threats to the security of the network and system. Intrusion detection system is one of the measure taken for the detection of the intrusion [50]. It mainly detects for the compromising of either data confidentiality, integrity, or availability. Based on the location of the intrusion detection system implementation, it is categorized into two types [20]; network based intrusion detection system and host based intrusion detection system. Network based intrusion detection system detects intrusion in the network segment, whereas host based intrusion detection detects intrusion in the host system. Despite the division of the intrusion detection system according to their implementation, intrusion detection system is further categorized according to the implementation of the detection method [54]. There are two types of intrusion detection method namely; signature based and anomaly based. Signature based detection method uses the stored signatures of the attacks for the detection of the intrusion. Due to the utilization of the stored signature of the attacks, this method has high true positive rate. This method, however, cannot detect zero day attack as no signature exist for zero day attack. On the other hand, anomaly based detection method can detect novel attacks as it works by taking the behavior of the network into consideration. Network anomaly detection method makes a baseline for the normal activity, any activity that deviates from that baseline is considered as a possible intrusion [38]. Anomaly based detection method, however, has a high false positive rate as it is difficult to map the normal behavior of the network. Introduction of new attacks in network changes the behavior of the network while the normal data behavior remains same. Anomaly detection is therefore, depends on the behavior of both normal data and anomalous data. Learning the boundary between normal behavior and anomalous behavior of the network is therefore required. In regard to separate the normal and anomalous behavior, many techniques including supervised classification algorithms are adapted for this purpose [14, 24]. These classification algorithms, however, highly depends on the input features of the data. Redundant, irrelevant, and noisy features make it difficult for the classification algorithm to build a detection model with high accuracy rate. Feature selection approach is therefore used, which selects the features that contribute more information about the class while not compromising the accuracy of the classification algorithm.

In network intrusion detection, features are extracted

from protocols header at different layers of network architecture and contents of data packets. Due to this reason noise in channels propagate to extracted features, this leads to false intrusion alarm. These noisy features should be removed using feature selection. In this work we used Ant colony optimization (ACO) for feature selection of the network data. ACO has optimal solution, because ACO can search in the feature space up to meeting an optimal solution. ACO has the historical linkage to previous iterations. Results of next iteration is based on the amount of pheromone, which is left in previous iterations. Every ant aimed to get the best local optimal solution and among those solution global optimal solution is found. The optimal feature set was then validated using Support Vector Machine (SVM) for classification of normal and intrusive activities in network.

The paper is organized as follows, Section 2 gives the brief literature review of the feature selection method for intrusion detection. Section 3 discusses the proposed feature selection algorithm and Section 4 discusses the results. The work is concluded in Section 5. .

2 Previous Work

Lin et al. [27] Used Simulated Annealing (SA) with Support Vector Machine (SVM) and Decision Tree (DT) together for feature selection and anomaly detection. Optimal feature set was selected using simulated annealing whereas Decision Tree was used to get rules from the dataset. SVM was used for classification of the data. Simulated annealing was also used to adjust automatically the parameter setting for SVM and DT. George [17] used Principle Component Analysis (PCA) along with SVM for intrusion detection. PCA was used for dimension reduction (feature selection) while SVM used for classification. Evaluation, based on the SVM with PCA approach, gave less misclassification compared to SVM method. Ganapathy et al. [16] along with survey of the different feature selection and classification algorithm, proposed new feature selection algorithm. This feature selection algorithm combined information gain ratio of the feature and rule based approach. Liu et al. [29] has used principle component analysis for feature selection along with neural networks for classification purpose. Features with highest eigenvalues were selected in the proposed approach. The mentioned approach, however, may leave the important features for inclusion. Solely relying on the high value of eigenvalues might not be enough for feature selection [4]. Baig et al. [5] proposed two-phase technique for the classification of KDD99 dataset into normal and anomalous class. This approach used three feature ranking techniques, gain ratio, information gain, and global method for data handling (GMDH) for feature selection.

Ghali [18] used the rough set theory along with the artificial neural network. The aim was to reduce the dataset for intrusion detection which resulted in less consumption of computer resources. RSNNA (Rough Set Neural Network Algorithm) was used for feature reduction, which found dependencies among the features while feed forward neural network was used for the classification of the data. Sheikhan et al. [45] proposed a method that used fuzzy association rule for the generation of feature subsets. While fuzzy logic with ARTMAP (adaptive resonance theory neural networks) was used for the validation of the feature subset using classification. Kannan et al. [23] proposed feature selection method based on genetic algorithm. The purpose of the study was to remove unimportant features thus reducing training time of the classification or clustering. In addition to that Fuzzy based SVM was used for the validation of the feature subset. The proposed genetic algorithm was based on weighted sum, increasing global search capability resulted in better attribute collaboration. Rufai et al. [42] combined membrane computing (MC) and bee algorithm (BA) for their work. Motivated by membrane structure and operations of living cells MC, gives the solution for BA to find the best feature subset. Thus, it improved BA for feature selection. BA was run on different membranes in the main membrane to get the initial solution. Zainal et al. [53] used a 2-tier approach, which included rough set and particle swarm optimization (PSO), Rough-PSO. SVM was used for classification while fitness function was used to find out the fitness of the proposed feature subset.

3 Dynamic Ant Colony System with Three Level Update Feature Selection (DACS3-FS)

Nature inspired science for solving many hard problems that are existed for humans. Researchers, therefore, mimics their properties to solve real world problems. One of them is following the foraging behavior of ants. Real ants have the property to solve very complex problems by utilizing information of each ant. Ants use a chemical substance called, pheromone, for indirect communication with each other. Pheromone is laid by ant on the way back from food to nest and vice versa. It works as a guidance to other ants. High pheromone intensity attracts more ants. Intensity of pheromone depicts the importance of the path. Using this property ants are able to select shortest path from nest to food source.

ACO was first used to solve Traveling Salesman Problem, which is a NP-complete problem [35]. Ant system is one of the variation of ant colony optimization technique. ACO is used for many optimization problems due to its high optimal solution for optimization problems [47]. Ant colony optimization (ACO) has less complexity in terms of time and memory requirement. ACO uses heuristic information and pheromone value to compute next move. During traversing an edge ant updates the pheromone value on the edges. Pheromone update also called global pheromone update for ant system. Ant system is one of the variation of ant colony optimization technique which

includes the pheromone update level after completion of the tour by all ants. This method was improved by introducing another level of pheromone update called; local pheromone update in ant colony system (ACS) [10].

Yi and Gong [25] introduced improved version of ACS called, Dynamic Ant Colony system (DACS). Improved version of ACS avoid the growth of pheromone level too high by introducing the dynamic decay parameter (1 - $\rho[\tau(r,s)]$). The dynamic decay parameter is applied at both level of pheromone updates such as local pheromone updating rule and global pheromone updating rule. Helmi et al. [39], improved the DACS algorithm by introducing updating of pheromone at three level; local level, intermediate level, and at global level. Local pheromone is updated when all ants start their tour. Intermediate pheromone updating is done by retrieving the best knowledge of the best individual ants of the group after completing a tour and then it is divided into best of the group and worst of all groups. This is followed by the global pheromone update in which the ant that having best tour is being considered and is divided into worst of the global best and best of the global best. This method provides better searching guidance in the effort to search for better solution.

In this research, the proposed work of Helmi et al. [39], is modified to adapt for the feature selection. Block diagram of the DACS3-FS algorithm is given in Figure 1. The purpose of declaring different pheromone update levels are to take the advantage of the pheromone intensity since the small increment in the pheromone values will not guarantee that ACO will give optimal solution [33]. Also by increasing amount of pheromone too high causes to converge the solution too early.

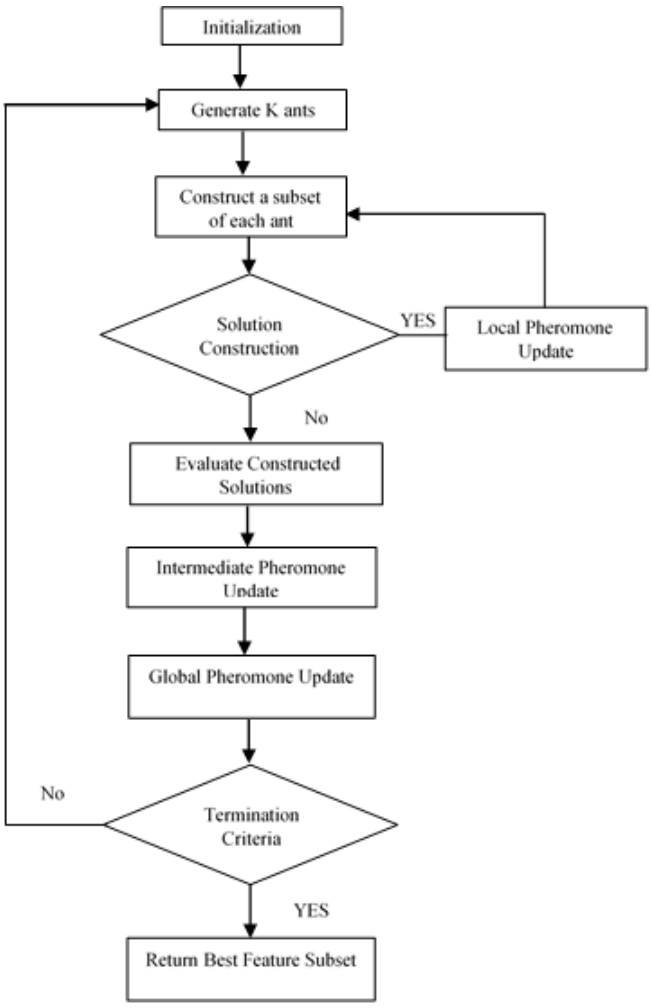

Figure 1: Flow chart of DACS3-FS

3.1 Features Representation

Suitable representation of the problem domain for ant colony optimization implementation is important. Some of the previous work used graphical representation for feature selection [40, 44], while some researchers used other methods like represented features in a binary form i.e. 1 and 0 [2, 46]. In this study, completely connected graph representation is used, Figure 2. Dark doted lines in the figure depicts, how features are selected by the best ant in feature selection process. Thus each ant have chance to select any node based on pheromone and heuristic value. Features are represented by a node and are connected with each other by edges. Pheromone and heuristic values are related with the features thus not laid on the arcs. Number of ants used were equal to total number of the features. Using too many ants would lead to quick convergence which may result to bad solution while few ants would not be able to utilize the cooperation of synergistic effect due to pheromone decay process [7].

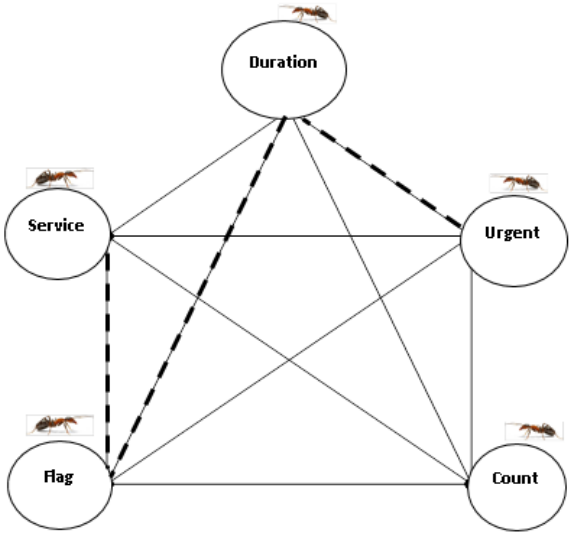

Figure 2: Graphical representation of features

3.2 Heuristic Function

Heuristic information plays a vital role to generate high quality results [37]. It describes priori desirability of the move. It has great influence on the performance of the ant colony optimization algorithms [28]. Artificial ants can lead to bad result and can reinforce it in each tour as initial random pheromones do not lead them. To avoid such deadlock heuristic information helps to recover from that deadlock [37]. Heuristic information related to the feature must be used [1]. It helps for positive constructive step thus helps to improve the performance of the ants [9]. There are two type of heuristic information [30] i.e. static heuristic and dynamic heuristic. Static heuristic is calculate and initialized at the start of the algorithm run and remains same throughout whole algorithms run like distance between cities in TSP. The advantage of the static heuristic is that it is calculated once and easy to compute [11]. Dynamic heuristic is calculated at every step as it depends on partial computed solution. Dynamic heuristic is, therefore, computationally expensive. In the proposed work we have used correlation as heuristic information. So the static heuristic is used in this work. Correlation of each feature respective to class is used and the values are constant thoughout whole algorithms run. So the ants can get some extra information for constructing solution. The importance of the heuristic is controlled by β.

3.3 Transition Probability

Like real ants artificial ants must evaluate the intensity of pheromone to do decision for next move [41]. In real ants the greater the amount of pheromone more the probability that ants will select that path. In computation problem completely relying on pheromone value for the path decision can lead to false result [36]. But in graph problem artificial ants decide the next move based on the probabilistic choice from a set of allowed nodes. This probabilistic choice depends on two parameters i.e. heuristic value and pheromone value [34]. Heuristic value is regarded as the visibility of the path in future perspective while pheromone value is regarded as common memory in past perspective of the path. Both values together controls the movement of the ant for solution constructions. In the proposed methodology we used correlation values of the features as a heuristic value and number of times features visited as initial pheromone value. Correlation values of the features to the classes will help move probability function to consider the importance of the feature to the prediction of the classes. In feature selection problem, the heuristic value should involve some kind of evaluation function for movement of feature to feature [21]. This will avoid ant to select bad feature each time by considering heuristic value since it is the priori desirability of the move. In the proposed method, initial pheromone value (τ_o)is set to 1 so no feature get biased pheromone at the start of the algorithm run. Thus allowing ant to select next feature unbiased regarding its pheromone value at start. Only feature will be selected once by a single ant while a single feature can be visited by many ants. Move probability for the proposed DACS3-FS is given as;

$$P_{ij} = max[(\tau_j)^{\alpha} * (\eta_j)^{\beta}] \quad (1)$$

Where τ_j represents pheromone value of the next feature while heuristic information of the next feature is represented by, η_j. Parameters α and β controls the trade-off between pheromone and heuristic information. The values of these parameters, therefore, influences the result of probability function. The trade-off between intensification and diversification is influenced by modifying the values of parameters [48]. Moreover, as static heuristic approach has been used in the proposed method, therefore, it will effect random probability function at the start of the construction. Selecting a single feature in each iteration, for feature subset construction, requires more computing cycles. Moreover, single feature by itself does not much help to find the class of the data, therefore, group of features are selected in this step.

3.4 Pheromone Value

Artificial pheromone is the cumulated numerical information associated with the edges that is laid by different ants during solution construction [12]. This pheromone information imitate the search experience of ants by changing the pheromone value by visiting edges each time. Pheromone information helps ant to probabilistically decide the next edge move. High intensity pheromone attracts more ants [8]. Pheromone evaporated with time that helps to diversify the search which improves the chances of other nodes to be explored by ants. Pheromone ants change this information each time they visit the edge to imitate and store their memory in the form of pheromone. This pheromone information is changed by different ants while traversing. Pheromone value, however, should be evaporated with some degree to diversify the traversing of the ants. At the end, each feature consists of the pheromone that has been laid by ants while traversing that feature. Feature which is visited by many ants will have high amount of pheromone.

3.4.1 Local Pheromone Update

Local search is the basic start of the ant colony optimization algorithm. Each arc in ant colony optimization get initialized by equally non-negative small value [51, 13]. An arc in ACO get local pheromone update with each iteration. Pheromones values get dwindle with each iteration. In ant colony system local pheromone is used to make visited edges less desirable and thus increase chance to explore the edges that are not visited yet [32, 19]. While in ant system we deposit the pheromone after decaying pheromone with some constant factor [49]. In this work local pheromone update, using Equation (2), is used while

ant traverse the feature.

$$\tau(r,s) \leftarrow (1-\rho).\tau(r,s) + \rho.\Delta\tau_o \quad (2)$$

Where

$$\Delta\tau_o = \begin{cases} \frac{1}{F_n} & if\ F_n\ \epsilon features\ visited\ by\ n\ ants \\ 0 & Otherwise \end{cases}$$

ρ is a pheromone decay parameter and $0<\rho<1$. Too much pheromone produced by the local pheromone update causes the local optimization solution in which it ignores the optimization solution [22]. To avoid local optimization solution problem the $\Delta\tau_o$ is controlled by the number of times the specific feature is used.

3.4.2 Intermediate Pheromone Update

In the proposed method we have used intermediate pheromone update which helps to reinforce the pheromone value. Available knowledge of each member of the ant′s group is used for intermediate pheromone update. Each ant′s subset was evaluated using naive bayes classifier. Recalling that feature subset selected by ant was evaluated using naive bayes classifier. Feature subset that produced high accuracy for naive bayes classifier was allowed to deposit pheromone using Equation (3). This step is important because it encourages ants to produce best feature subset for global pheromone update level.

$$\tau(r,s) \leftarrow ((1-\rho.\tau(r,s)).\tau(r,s)) + \rho.\Delta\tau_o \quad (3)$$

Where

$$\Delta\tau_o = \begin{cases} \frac{1}{accuracy} & if\tau(r,s)\epsilon Local\ Group\ Best\ Tour \\ -\frac{1}{accuracy} & if\tau(r,s)\epsilon Local\ Group\ Worst\ Tour \\ 0 & Otherwise \end{cases}$$

3.4.3 Global Pheromone Update

The accuracy from one classification algorithm is not enough. This is because one classification algorithm does not necessarily be able to find the correct relationship between the feature and class label. As pheromone update rule does not necessarily produce best solution [33]. Similarly, subset that produced good accuracy in intermediate pheromone update level is not necessary be able to produce good accuracy result for other classification algorithms. Another level of pheromone deposit, therefore, helps towards finding of best solution. Global pheromone also help to tackle the local optima problem. In the proposed ACO algorithm, accuracy of Support Vector Machine (SVM) was used for the global pheromone update. Refer to the intermediate pheromone update level, only a single ant is able to deposit the pheromone which is selected feature set produced high accuracy for SVM. Best of the best i.e. the subset which was able to produce high accuracy for both classification algorithm can deposit the global pheromone using Equation (4).

$$\tau(r,s) \leftarrow ((1-\rho.\tau(r,s)).\tau(r,s)) + \rho.\Delta\tau_o \quad (4)$$

Where

$$\Delta\tau_o = \begin{cases} \frac{1}{accuracy} & if\tau(r,s)\epsilon Global\ Group\ Best\ Tour \\ -\frac{1}{accuracy} & if\tau(r,s)\epsilon Global\ Group\ Worst\ Tour \\ 0 & Otherwise \end{cases}$$

3.4.4 Stopping Criteria

ACO runs several times until some stopping criteria met. Stopping criteria can be number of time algorithms run, number of subsets evaluation, maximum number of iterations, or for specific number of time the best solution is not changed and so on [31]. In this work maximum number of iterations were used as stopping criteria. Apart from that if any ant is unable to improve accuracy in three runs than this ant is destroy. This is necessary to avoid trap into local optima.

4 Results and Discussions

KDD99 dataset is the benchmark dataset used for the evaluation of anomaly detection methods in network intrusions [43]. Many research groups validated their detection model using KDD99 dataset [3, 6, 26, 52]. The dataset came from DARPA98 IDS evaluation program [15]. Training data is collected from seven weeks of data in which few weeks data are attack free while other weeks of data consist of attacks. Despite of it, two weeks of data resulted testing dataset which consist of attack data and normal data. Kdd99 has huge records thaś why its subset is widely used and is called kddcup.data_10_percent (kdd99_10%). 22 attacks are in the training set, 16 additional attacks are in the testing set. The training set contains 494020 instances while 311029 instances are for testing dataset. KDD99 contains four attack classes, User-to-Root (U2R), Probe, Denial of Service (DoS), Root-to-Local (R2L), and one legitimate data class called, Normal.

The experiments were carried out on system with core i7 and running windows 10 with 16GB of RAM. Matlab and weka tools were used for the experiment. Moreover, KDD99 dataset contains redundant data which were removed in this experiment. Support Vector Machine (SVM) was used for the validation of the feature subset selected using DACS3-FS. SVM was used for both binary and multiclass classification. Results of the feature subsets are empirically compared with the benchmark results.

The features that have been used in previous studies are shown in Table 1. As discussed earlier KDD99 dataset has 41 features thus we also used the result of the KDD99 full feature set for comparison purpose. This is due to the purpose of the study to increase true positive rate (TPR), when the data is correctly classified in its own class, and precision, the portion of the true positive over all the positive instances the detection method has detected as anomalous, and accuracy meanwhile minimizing false positive rate (FPR), when data of some other

class is incorrectly accepted, as low as possible. The F-measure is the harmonic mean of precision and recall. A good classifier is expected to obtain F-measure as high as possible. These features were validated for both binary and multiclass classification.

Table 1: Features selected using different feature selection methods

FS Method	Given Features	Authors
Information Gain	2,5,8,10,14,15,19, 26,27,30,31,32,33, 34,35,36,37,38,40	Ganapathy et al[16]
Rough Set	5,6,23,24,32, 33,36	Ghali [18]
Genetic Algorithm	2,3,4,5,6,10,12,23, 25,29,30,35, 36, 37,38,40	Kannan et al.[23]
Membrane Computing	2,3,8,13,20,24,32, 37,37,39,40	Rufai et al.[42]
KDD99	41	—
DACS3-FS	2,3,5,6,23, 33	—

4.1 Binary Classification

These features were validated for both binary and multiclass classification. As discussed in earlier section, KDD99 dataset has one legitimate class called, Normal, and four attack classes, DoS, R2L, PROBE, and U2R. For binary classification, these four attack classes were combined into a single class i.e. Attack class.

Table 2 gives the comparison of the different feature selection method. As shown in table, feature set selected using DACS3-FS algorithm had out performed resulted accuracy 98.7087% while full feature set resulted accuracy of 98.5172%. Moreover, Table 3 gives the detail comparison for normal class result, it can be seen that DACS3-FS based feature set had performed well for binary classification. It had though TPR 99.1% slightly less than the TPR 99.2% of rough set based feature set but FPR for DACS3-FS was less compared to other feature sets results. Result of attack class for different feature set approach is given in Table 4. Rough set based feature set had FPR and precision slightly better than DACS3-FS based feature set but the classification algorithm was unable to classify attack class data.

Table 2: Binary accuracy comparison

FeatureAlgorithms	Features	Accuracy%
IG	19	97.6348
Rough Set	7	98.0191
MC	10	95.9747
GA	17	98.3645
KDD99	41	98.5172
DACS3-FS	6	**98.7087**

Table 3: Normal class result comparison

Algorithm	TPR %	FPR %	Precision %	F-measure
Information Gain	98.9	4.3	97.2	0.981
Rough Set	99.2	3.8	97.6	0.984
Membrane Computing	99.1	8.9	94.5	0.968
Genetic Algorithm	99.1	2.7	98.3	0.987
DACS3-FS	99.1	1.9	98.8	0.989

Table 4: Attack class result comparison

Algorithm	TPR %	FPR %	Precision %	F-measure
Information Gain	95.7	1.1	98.3	0.969
Rough Set	96.2	0.8	98.7	0.974
Membrane Computing	91.1	0.9	98.5	0.947
Genetic Algorithm	97.3	0.9	98.5	0.979
DACS3-FS	98.1	0.9	98.6	0.984

4.2 Multiclass Classification

In the previous section, we used binary SVM for the binary classification of KDD99 dataset, in which we combined all attack classes into a single attack class. In this section, result for multiclass classification is given. SVM was used for multiclass, although it is a binary classifier but it can be used for multiclass using cascading different binary SVM.

For normal class result is given in the Table 5. DACS3-FS based feature set resulted low FPR and high precision compared to other feature sets. For DoS attack class different feature sets result is given in Table 6. It can be seen the genetic algorithm based feature set had better result compared to DACS3-FS based feature set and information gain based feature set. TPR for these three feature set were same.

DACS3-FS based feature set performed well for R2L class as given in Table 7. It had high TPR 80% and FPR 0%. Rough set based feature set had high precision and low FPR but had low TPR 60.6% for Probe class as given in Table 8. U2R class result using different feature sets is given in Table 9. It can be rough set and DACS3-FS and rough set based feature sets had same TPR 52.2% and FPR 0.0% but DACS3-FS based feature set resulted high precision. Accuracy for different feature selection algorithm is given in Table 10, for different classes. It can be seen that DACS3-FS resulted high accuracy of 98.7359% while full feature set resulted accuracy of 98.6013%.

Table 5: Normal class result comparison

Algorithm	TPR %	FPR %	Precision %	F-measure
Information Gain	99.5	5.1	96.8	0.981
Rough Set	99.4	3.3	97.9	0.987
Membrane Computing	99.4	8.5	94.7	0.970
Genetic Algorithm	99.7	3.2	98.0	0.988
DACS3-FS	99.5	1.5	99.1	0.993

Table 6: DoS class result comparison

Algorithm	TPR%	FPR%	Precision%	F-measure
Information Gain	99.4	0.6	99.0	0.992
Rough Set	99.6	1.0	98.2	0.989
Membrane Computing	96.6	2.6	95.3	0.959
Genetic Algorithm	99.6	0.4	99.2	0.994
DACS3-FS	99.6	0.7	98.7	0.992

Table 7: R2L class result comparison

Algorithm	TPR %	FPR %	Precision %	F-measure
Information Gain	17.9	0.0	98.7	0.303
Rough Set	55.7	0.1	91.7	0.693
Membrane Computing	1.5	0.1	15.0	0.028
Genetic Algorithm	68.4	0.0	95.8	0.798
DACS3-FS	80.9	0.0	97.0	0.882

Table 8: PROBE class result comparison

Algorithm	TPR %	FPR %	Precision %	F-measure
Information Gain	81.1	0.2	91.9	0.861
Rough Set	60.6	0.1	92.2	0.731
Membrane Computing	15.4	0.1	84.1	0.260
Genetic Algorithm	71.6	0.2	90.5	0.800
DACS3-FS	80.8	0.2	90.4	0.853

Table 9: U2R class result comparison

Algorithm	TPR %	FPR %	Precision %	F-measure
Information Gain	0.0	0.0	0.0	0.000
Rough Set	52.2	0.0	80.7	0.634
Membrane Computing	0.0	0.0	0.0	0.000
Genetic Algorithm	3.8	0.0	57.1	0.072
DACS3-FS	52.2	0.0	90.1	0.661

Table 10: Multiclass accuracy comparison

Feature Algorithms	Features	Accuracy%
IG	19	97.4769
Rough Set	7	97.834
MC	10	94.7481
GA	17	98.2571
DACS3-FS	6	**98.7359**

5 Conclusion

The performance of the classification algorithms is highly depends on input features of the data. Poor selection of features can affect classification accuracy badly which leads toward high rates of false negatives and false positives. This problem can be handled effectively by using optimized selected features. Many feature selection methods are unable to identify the complex relationship between the features which in result unable to produce the useful features. Some ranking feature selection methods tried to find all the features relevant to class attribute but failed to identify the redundant features. In this study novel feature selection algorithm called, Dynamic Ant Colony System with Three Level Update Feature Selection, a variant of ant colony optimization was proposed. The proposed algorithm is a wrapper based feature selection approach using two machine learning algorithm for the evaluation of the feature set during feature selection process. The proposed feature selection algorithm resulted in an optimal feature set that produced efficient detection model in terms of accuracy compared to the previous feature selection algorithms.

References

[1] N. Abd-Alsabour, H. Hefny, and A. Moneim, "Heuristic information for ant colony optimization for the feature selection problem," in *Conference Anthology*, pp. 1–5, IEEE, 2013.

[2] N. Abd-Alsabour and M. Randall, "Feature selection for classification using an ant colony system," in *Sixth IEEE International Conference on e-Science Workshops*, pp. 86–91, 2010.

[3] A. Abraham, C. Grosan, and C. Martin-Vide, "Evolutionary design of intrusion detection programs.," *International Journal of Network Security*, vol. 4, no. 3, pp. 328–339, 2007.

[4] I. Ahmad, M. Hussain, A. Alghamdi, and A. Alelaiwi, "Enhancing SVM performance in intrusion detection using optimal feature subset selection based on genetic principal components," *Neural Computing and Applications*, vol. 24, no. 7-8, pp. 1671–1682, 2014.

[5] Z. A. Baig, S. M. Sait, and A. Shaheen, "Gmdh-based networks for intelligent intrusion detection," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 7, pp. 1731–1740, 2013.

[6] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Towards generating real-life datasets for network intrusion detection.," *International Journal of Network Security*, vol. 17, no. 6, pp. 683–701, 2015.

[7] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm intelligence: from natural to artificial systems*, No. 1, Oxford university press, 1999.

[8] M. Ciba and I. Sekaj, "Ant colony optimization with re-initialization," *Automation, Control and Intelligent Systems*, vol. 1, no. 3, pp. 59–63, 2013.

[9] C. R. Conti, M. Roisenberg, and G. S. Neto, "ACO - v-an algorithm that incorporates the visibility heuristic to the aco in continuous domain," in *IEEE Congress on Evolutionary Computation*, pp. 1–8, 2012.

[10] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 28–39, 2006.

[11] M. Dorigo and T. Stützle, "The ant colony optimization metaheuristic: Algorithms, applications, and advances," in *Handbook of Metaheuristics*, pp. 250–285, Springer, 2003.

[12] M. Dorigo and T. Stützle, "Ant colony optimization: overview and recent advances," in *Handbook of Metaheuristics*, pp. 227–263, Springer, 2010.

[13] D. D. Duc, H. Q. Dinh, and H. H. Xuan, "On the pheromone update rules of ant colony optimization approaches for the job shop scheduling problem," in *Pacific Rim International Conference on Multi-Agents*, pp. 153–160, Springer, 2008.

[14] M. A. Eid, H. Artail, A. I. Kayssi, and A. Chehab, "Lamaids: A lightweight adaptive mobile agent-based intrusion detection system.," *International Journal of Network Security*, vol. 6, no. 2, pp. 145–157, 2008.

[15] M. A. Faisal, Z. Aung, J. R. Williams, and A. Sanchez, "Data-stream-based intrusion detection system for advanced metering infrastructure in smart grid: A feasibility study," *IEEE Systems Journal*, vol. 9, no. 1, pp. 31–44, 2015.

[16] S. Ganapathy, K. Kulothungan, S. Muthurajkumar, M. Vijayalakshmi, P. Yogesh, and A. Kannan, "Intelligent feature selection and classification techniques for intrusion detection in networks: A survey," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, p. 1, 2013.

[17] A. George, "Anomaly detection based on machine learning: dimensionality reduction using pca and classification using SVM," *International Journal of Computer Applications*, vol. 47, no. 21, 2012.

[18] N. I. Ghali, "Feature selection for effective anomaly-based intrusion detection," *International Journal of Computer Science and Network Security*, vol. 9, no. 3, pp. 285–289, 2009.

[19] S. Gilmour and M. Dras, "Understanding the pheromone system within ant colony optimization," in *Australasian Joint Conference on Artificial Intelligence*, pp. 786–789, Springer, 2005.

[20] G. Javadzadeh and R. Azmi, "Idufg: Introducing an intrusion detection using hybrid fuzzy genetic approach.," *International Journal of Network Security*, vol. 17, no. 6, pp. 754–770, 2015.

[21] R. Jensen, *Combining Rough and Fuzzy Sets for Feature Selection*, PhD thesis, Citeseer, 2005.

[22] X. JunYong, H. Xiang, L. CaiYun, and C. Zhong, "A novel parallel ant colony optimization algorithm with dynamic transition probability," in *International Forum on Computer Science-Technology and Applications (IFCSTA'09)*, vol. 2, pp. 191–194, 2009.

[23] A. Kannan, G. Q. Maguire Jr, A. Sharma, and P. Schoo, "Genetic algorithm based feature selection algorithm for effective intrusion detection in cloud networks," in *2012 IEEE 12th International Conference on Data Mining Workshops*, pp. 416–423, 2012.

[24] F. Kuang, W. Xu, and S. Zhang, "A novel hybrid kpca and SVM with ga model for intrusion detection," *Applied Soft Computing*, vol. 18, pp. 178–184, 2014.

[25] Y. Li and S. Gong, "Dynamic ant colony optimisation for tsp," *The International Journal of Advanced Manufacturing Technology*, vol. 22, no. 7-8, pp. 528–533, 2003.

[26] Z. Li and A. Das, "The utility of partial knowledge in behavior models: An evaluation for intrusion detection," *International Journal of Network Security*, vol. 1, no. 3, pp. 138–146, 2005.

[27] S. Lin, K. Ying, C. Lee, and Z. Lee, "An intelligent algorithm with feature selection and decision rules applied to anomaly intrusion detection," *Applied Soft Computing*, vol. 12, no. 10, pp. 3285–3290, 2012.

[28] Y. Lin and J. Zhang, "Ant colony optimization with adaptive heuristics design," in *Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation*, pp. 3–4, ACM, 2013.

[29] G. Liu, Z. Yi, and S. Yang, "A hierarchical intrusion detection model based on the pca neural networks," *Neurocomputing*, vol. 70, no. 7, pp. 1561–1568, 2007.

[30] M. J. Meena, K. R. Chandran, and J. M. Brinda, "Integrating swarm intelligence and statistical data for feature selection in text categorization," *Pheromones*, vol. 1, p. 15, 2010.

[31] D. Merkle and M. Middendorf, "Swarm intelligence," in *Search Methodologies*, pp. 213–242, Springer, 2014.

[32] R. Montemanni, L. M. Gambardella, A. E. Rizzoli, and A. V. Donati, "Ant colony system for a dynamic vehicle routing problem," *Journal of Combinatorial Optimization*, vol. 10, no. 4, pp. 327–343, 2005.

[33] A. Moraglio, F. E. Otero, and C. G. Johnson, "The ACO encoding," in *International Conference on Swarm Intelligence*, pp. 528–535, Springer, 2010.

[34] L. Nunes de Castro, "Fundamentals of natural computing: an overview," *Physics of Life Reviews*, vol. 4, no. 1, pp. 1–36, 2007.

[35] Z. A. Othman, H. M. Rais, and A. R. Hamdan, "Embedding malaysian house red ant behavior into an ant colony system," *Journal of Computer Science*, vol. 4, no. 11, p. 934, 2008.

[36] W. Pan and L. Wang, "An ant colony optimization algorithm based on the experience model," in *2009 Fifth International Conference on Natural Computation*, vol. 3, pp. 13–18, 2009.

[37] S. Parsons, *Ant Colony Optimization by Marco Dorigo and Thomas StÜTzle*, MIT Press, ISBN 0-262-04219-3, 2005.

[38] Q. S. Qassim, A. M. Zin, and M. J. Ab Aziz, "Anomalies classification approach for network-based intrusion detection system," *International Journal of Network Security*, vol. 18, no. 6, pp. 1159–1172, 2016.

[39] H. M. Rais, Z. A. Othman, and A. R. Hamdan, "Improved dynamic ant colony system (dacs) on symmetric traveling salesman problem (tsp)," in *International Conference on Intelligent and Advanced Systems (ICIAS'07)*, pp. 43–48, 2007.

[40] M. H. Rasmy, M. El-Beltagy, M. Saleh, and B. Mostafa, "A hybridized approach for feature selection using ant colony optimization and ant-miner for classification," in *8th International Conference on Informatics and Systems (INFOS'12)*, pp. BIO–211, 2012.

[41] I. J. Riadi, *Cognitive Ant Colony Optimization: A New Framework In Swarm Intelligence*, PhD thesis, University of Salford, 2014.

[42] K. I. Rufai, R. C. Muniyandi, and Z. A. Othman, "Improving bee algorithm based feature selection in intrusion detection system using membrane computing," *Journal of Networks*, vol. 9, no. 3, pp. 523–529, 2014.

[43] C. Scheper and W. J. Roberts, "Anomaly detection using an mmpp-based glrt.," *International Journal of Network Security*, vol. 17, no. 6, pp. 672–677, 2015.

[44] W. Shahzad, *Classification and Associative Classification Rule Discovery Using Ant Colony Optimization*, PhD thesis, National University of Computer & Emerging Sciences, 2010.

[45] M. Sheikhan, M. S. Rad, and H. M. Shirazi, "Application of fuzzy association rules-based feature selection and fuzzy artmap to intrusion detection," *Majlesi Journal of Electrical Engineering*, vol. 5, no. 4, 2011.

[46] Q. Shen, J. Jiang, J. Tao, G. Shen, and R. Yu, "Modified ant colony optimization algorithm for variable selection in qsar modeling: Qsar studies of cyclooxygenase inhibitors," *Journal of Chemical Information And Modeling*, vol. 45, no. 4, pp. 1024–1029, 2005.

[47] R. K. Sivagaminathan and S. Ramakrishnan, "A hybrid approach for feature subset selection using neural networks and ant colony optimization," *Expert Systems with Applications*, vol. 33, no. 1, pp. 49–60, 2007.

[48] C. Solnon and D. Bridge, "An ant colony optimization meta-heuristic for subset selection problems," *System Engineering using Particle Swarm Optimization, Nova Science*, vol. 729, 2006.

[49] T. Stützle and M. Dorigo, "ACO algorithms for the traveling salesman problem," *Evolutionary Algorithms in Engineering and Computer Science*, pp. 163–183, 1999.

[50] M. Uddin, A. A. Rahman, N. Uddin, J. Memon, R. A. Alsaqour, and S. Kazi, "Signature-based multilayer distributed intrusion detection system using mobile agents.," *International Journal of Network Security*, vol. 15, no. 2, pp. 97–105, 2013.

[51] J. Van Ast, R. Babuška, and B. De Schutter, "Generalized pheromone update for ant colony learning in continuous state spaces," in *IEEE Congress on Evolutionary Computation*, pp. 1–8, 2010.

[52] W. Wang, X. Zhang, S. Gombault, and S. J Knapskog, "Attribute normalization in network intrusion detection," in *10th International Symposium on Pervasive Systems, Algorithms, and Networks*, pp. 448–453, 2009.

[53] A. Zainal, M. A. Maarof, and S. M. Shamsuddin, "Feature selection using rough-dpso in anomaly intrusion detection," in *International Conference on Computational Science and Its Applications*, pp. 512–524, Springer, 2007.

[54] J. Zeng and D. Guo, "Agent-based intrusion detection for network-based application.," *International Journal of Network Security*, vol. 8, no. 3, pp. 201–210, 2009.

Helmi Md Rais is a senior lecturer at Universiti Teknologi PETRONAS (UTP) under the Faculty of Science and Information Technology (FSIT). He received his PhD degree in Science and System Management in 2013 from Universiti Kebangsaan Malaysia (UKM), Malaysia. He received his BSc degree from Drexel University, USA in 1999 and his Master degree from Griffith University, Australia in 2001. His research interests include swarm intelligent, database and management information systems.

Tahir Mehmood received his BS in Computer Science from University of Peshawar, Pakistan. He is currently pursuing MS in Information Technology from Universiti Teknologi PETRONAS, Malaysia. His research interests include machine learning, big data, image processing, data science, and network security.