Ant Colony Optimization and Feature Selection for Intrusion Detection

Tahir Mehmod and Helmi B. Md Rais

Abstract Network intrusion detection gained a lot of attention from the security expert. Intrusion detection system has been designed for the purpose detecting attack and comprises of detection method that can be anomaly based or it can be signature based. These detection method, however, highly depends on the quality of the input features. Supervised learning approach for the detection method finds the relationship between the feature and its class. Therefore, irrelevant, redundant, and noisy features must be eliminated before applying supervised algorithm. This can be done by feature selection method. In this paper ant colony optimization has been applied for feature selection on KDD99 dataset. The reduced dataset is validated using support vector machine. Results show that accuracy of the SVM is significantly improved with reduced feature set.

1 Introduction

As many organizations are facing the cyber-attack, confidentiality, integrity, and availability of the data is become a major issue. Intrusion detection system is designed to detect intrusion in a single host or in a network [1]. Formal type is called host based intrusion detection system while the later one is the second type of intrusion detection system and is called, network based intrusion detection system. Host based intrusion detection system use system log files and other logging mechanism to identify any attack. It resides on a single host system and that is why it highly depends on operating system architecture. Any shortcoming of operating system may compromise intrusion detection system as well. On the other hand, network based intrusion detection system is deployed on a network segment and it

T. Mehmod (✉) · H.B.M. Rais
Department of Computer and Information Sciences,
Universiti Teknologi Petronas, Seri Iskandar, Malaysia
e-mail: tahirmehmood.seecs@gmail.com

H.B.M. Rais
e-mail: helmim@petronas.com

P.J. Soh et al. (eds.), *Advances in Machine Learning and Signal Processing*,
Lecture Notes in Electrical Engineering 387, DOI 10.1007/978-3-319-32213-1_27

analyzes the network packet for any attack [2]. Network based intrusion detection system is independent of the operating system, in fact it is transparent to the operating system of the host as it do not reside on the host system. Both types of intrusion detection system use intrusion detection method for the detection of intrusion.

There are two types of intrusion detection method. One is called signature based intrusion detection method and the second type is called anomaly based intrusion detection method [3]. Signature based intrusion detection method also known as misuse based detection method looks for pattern or signature in a data that complies within a malware [4]. Signature based IDS has a database consists of signatures of the attacks. Signature based IDS uses the stored signatures of the malware for the detection. Therefore, this method has high true positive rate. The problem with the signature based technique is that it cannot detect novel attacks as no signature exist yet for the novel attack [5]. Contrarily anomaly based detection method can detect novel attack as it looks for abnormal behavior that do not comply with the normal operation of the system or network. Major drawback of this method is that it has high false positive rate since it is difficult to define normal behavior of the system or network [6].

Many machine learning algorithms have been used for the implementation of the detection methods. These machine learning algorithms highly depend on the input features. Irrelevant, redundant, and noisy features causes the machine learning algorithm to develop the detection model with low accuracy rate and with high false positive rate. Therefore, these feature must be eliminated at the preprocessing step.

Rest of the report is organized as follows; Sect. 2 gives introduction about feature selection. Section 3 contains introduction of ant colony optimization (ACO) which is followed by related work in Sect. 4. Proposed methodology has been discussed in Sect. 5 and results are given and discussed in Sect. 6. At the end conclusion is given in Sect. 7.

2 Feature Selection

For classification problem feature selection is used for the elimination of irrelevant, redundant, and noisy features to improve the accuracy of the classification algorithm. This is done at the preprocessing step before applying any machine learning algorithm. Feature selection process selects a subset of features that represents the whole feature set [7]. Which features should be included or excluded is being decided in this process. Relevancy and redundancy are the two decisive factors for feature selection process [8]. Relevant features are those that envisage the desired system response, on the other hand, redundant features have a high degree of correlation among themselves. Thus, removal of the redundant features is desired. Features that are highly correlated with each other give no additional information. While robust features have a high degree of correlation relevant to desired decision and uncorrelated with other features in feature set. By using feature selection at

preprocessing step, the predictive accuracy of the machine learning algorithms can be increased. Robust feature set also reduces the training time of the classifier as robust features are invariant in nature and reduces the dimension in high dimensional data. Reduced dataset also decreases dataset which acquire less storage space.

Feature selection process has four steps as shown in Fig. 1. Subset generation, subset evaluation, stopping criteria, and result validation [9]. Subset generation generates a different subset of features, and each feature subset is evaluated in subset evaluation process. If the current feature subset is better than previous feature subset than it replaces the previous one. Subset generation is a searching process, which can be complete, heuristic or random search. Generated feature subset is then validated by some tests.

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. There are two types of feature selection methods: Filter and wrapper. Filter method selects the subset of features without involving learning algorithm in evaluation phase and is mainly based on ranking of features, which represents the relevancy of the features [10]. In contrast, wrapper method evaluates a subset of features using learning algorithm [11]. This evaluating algorithm is called iteratively unless a robust subset of the feature is selected. Filter based approach is computationally fast compared to wrapper based as it doesn't involve any learning algorithm during ranking of features. However, wrapper based feature subset accomplishes good accuracy rate as it involves learning algorithm in the subset evaluation phase [12].

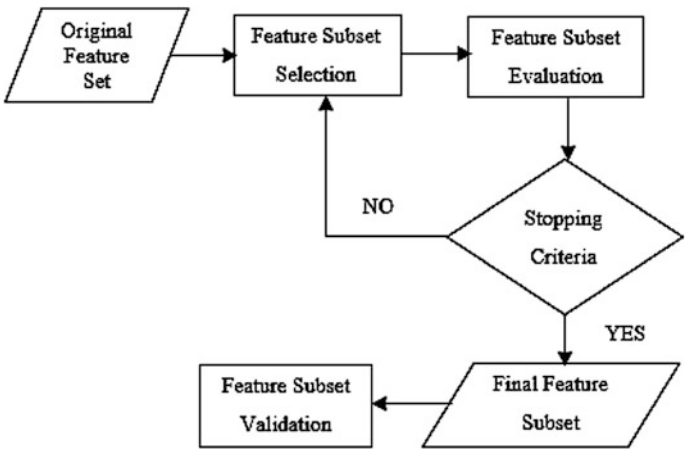

Fig. 1 Feature selection process

3 Ant Colony Optimization

Ants used chemical substance for the communication and is called, pheromone. Ants used it to remember the path from source of food to nest. More intensity of pheromone attracts more ants. Initially ant's foraging like behavior was used for traveling salesman problem (TSP) and the model was named ant system [13]. Ant colony optimization (ACO) has produced efficient results for NP-hard set problems. ACO has many variations. In this paper we used ant colony system (ACS), which uses two level pheromone update i.e. local pheromone and global pheromone update. In global pheromone update only those edges get pheromone update that belongs to best ant. A digital ant selects next node using some transition probability rule.

4 Related Work

Since feature selection is NP-complete problem that is why ACO has been widely adapted for feature selection. Below lists some of them.

George [14] utilized principle component analysis for feature reduction. Using principle component analysis 28 features were selected. The reduced feature set was validated using SVM. Tsang et al. [15] used independent component analysis and principle component analysis for his proposed model called, ant colony clustering model.

Gao et al. [16] proposed ant colony optimization method for KDD99 feature selection. The proposed method mapped the features into graph which were connected to each other, giving opportunity to each ant to select any feature. Selection of next feature by an ant was based on the heuristic information and pheromone value. Fisher discrimination rate was used as heuristic information. Edges contained the pheromone value and only the ant that resulted less squared error used global pheromone update on the edges visited during solution construction.

Nadia and Marcus [17] proposed a wrapper based feature selection based on ACO. The proposed method does not used traditional graph method, instead each feature is represented by 1 and 0 which indicated the selection of feature. An ant select next feature using a probability function which uses pheromone value and heuristic information. Each feature possessed pheromone value. While heuristic information described the desirability of feature which calculated by the number of ants visited that feature. Local pheromone update was used at each construction step while global pheromone was used by the best ant which update the pheromone value for all features that were selected by best ant.

Alwan and Mahamud [18] used mixed variable ant colony optimization for feature selection and at the mean time regulating C and γ parameters for SVM. SVM used RBF kernel function and the applied kernel function highly depends on C and γ value. During the feature selection mixed variable ant colony optimization method also searches for C and γ values that can improve the accuracy of SVM.

Heuristic information adopted fisher discrimination rate. Pheromone value lied on the edges and only the ant that produced high accuracy for SVM was allowed to update the pheromone value on the edges used during solution construction.

5 Proposed Methodology

Ant colony optimization for feature selection has been proposed in this paper. Features are represented in a completely connected graph problem thus choice of selecting next feature is given to each ant. An ant moves to next feature using given transition probability.

$$p_{i,j} = \max(\tau_j)^{\beta}.\eta_j \tag{1}$$

Pheromone value (τ) is related to each feature instead on edges. At the start of solution pheromone value to each feature is initialized by its entropy value to the prediction of the class. β controls the importance of the pheromone value during selection of next feature. If β is 0 than pheromone value for the feature is completely ignored. Number of time the feature visited is considered as heuristic information (η). Initially heuristic information is kept 1 so that no feature can get biased heuristic at the start of the constructing solution by the ants.

At each solution construction, local pheromone value is updated using given formula,

$$\tau_j = (1-\rho).\tau_j + \triangle_j \cdot \sigma \tag{2}$$

where

$$\Delta_j = \begin{cases} \rho & if\ j\ \varepsilon\ S^+ \\ 0 & otherwise \end{cases}$$

S^+ is the set of features visited for that particular run. $\sigma \in (0,1]$ controls the value for Δ_j. After completing tour each ant passes its dataset to naïve bayes classifier. Ant's dataset that results high accuracy rate gets global pheromone update and uses following equation.

$$\tau_j = (1-\rho).\tau_j + \emptyset_j \cdot \sigma \tag{3}$$

Where

$$\emptyset_j = \begin{cases} \rho & if\ j\ \varepsilon\ global\ best\ ant\ tour \\ 0 & otherwise \end{cases}$$

This whole method is repeated until stop criteria is met which is if none of the ants can improve accuracy for naive bayes classifier compare to the previous best ant result.

6 Results and Discusssions

In this paper KDD99 dataset has been used for evaluation of detection model. Feature subset is validated using LibSVM in Weka [19]. Binary classification has been used in the experiment. This dataset is widely adapted for the evaluation of detection model [20]. Dataset contains one normal class data and four attack classes' data namely, Denial of service (DoS), Probe, Remote-to-Local (R2L), and User-to-Root (U2R). Training and testing dataset exist for this dataset. Training dataset hold 494,021 network records while testing dataset comprises of 311,029 network records. Each instance is represented by 41 features. Both datasets contains redundant instances which were removed. Subset of training dataset is generated which contains 5823 instances for each two classes, normal class and attack class. Since we used binary SVM therefore all the four attack classes are merged into single attack class. The test dataset used contains 77,287 instances.

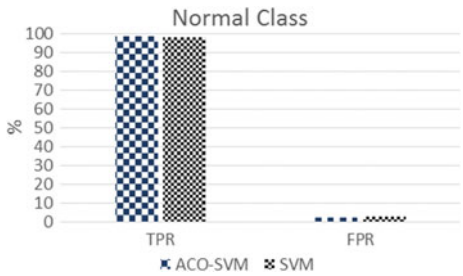

Fig. 2 Result comparison for normal class

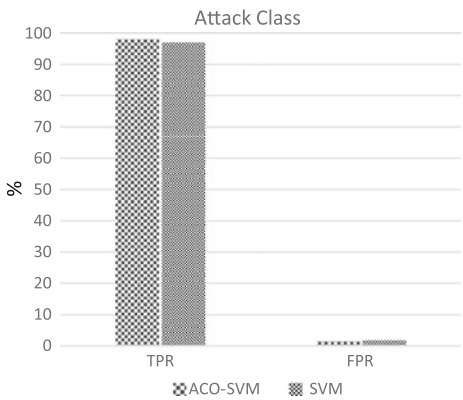

Fig. 3 Result comparison for attack class

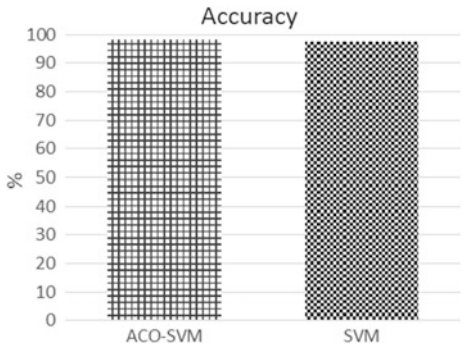

Fig. 4 Accuracy comparison for feature set

By using ACO 14 features were selected. Feature subset is than validated using SVM. Result is compared with full feature set result. Figure 1 shows the result for normal class. It can be seen that true positive rate (TPR) for reduced feature set is improve to 98.5 % from 98.2 % for whole feature set. Moreover, false positive rate (FPR) for reduced feature set is 2 % compared to full feature set i.e. 3 %. Figure 2 depicts the result for the attack class. From result it can be seen that reduced feature set gave TPR 98 % which is better than full feature set results 97 %. Accuracy for both feature set is shown in Fig. 3. Accuracy rate for SVM has been improved from 97.72 to 98.29 % when classified with the reduced feature set. (See Fig. 4)

7 Conclusion

High amount of data and irrelevant, redundant features make it difficult to build the prediction model for anomaly detection method. Features selection play vital role to build the prediction model in machine learning. In this paper feature selection method for anomaly detection has been presented. Ant colony optimization (ACO) has been proposed in the work due to its capability of utilizing previous information in the form of pheromones. SVM is used to build the anomaly detection model and the selected features are validated using this model. This work shows that robust features can be selected using ACO. Also fast and efficient detection method can be achieved using these robust features. This leads to real time detection of the intrusions in networks.

References

1. Kenkre PS, Pai A, Colaco L (2015) Real time intrusion detection and prevention system. In: Proceedings of the 3rd international conference on frontiers of intelligent computing: theory and applications (FICTA) 2014, pp 405–411

2. Bhuyan MH, Bhattacharyya DK, Kalita JK (2014) Network anomaly detection: methods, systems and tools. IEEE Commun Surv Tutorials 16(1):303–336
3. Othman ZA, Muda Z, Theng LM, Othman MR (2014) Record to record feature selection algorithm for network intrusion detection. Int J Adv Comput Technol 6(2):163
4. García-Teodoro P, Díaz-Verdejo J, Maciá-Fernández G, Vázquez E (2009) Anomaly-based network intrusion detection: techniques, systems and challenges. Comput Secur 28(1–2):18–28
5. Hämäläinen T (2014) Artificial immune system based intrusion detection: innate immunity using an unsupervised learning approach
6. Friedberg I, Skopik F, Fiedler R (2015) Cyber situational awareness through network anomaly detection: state of the art and new approaches. e i Elektrotechnik und Informationstechnik 132 (2):101–105
7. García S, Luengo J, Herrera F (2015) Feature selection. In: Data preprocessing in data mining SE—7, vol 72. Springer, pp 163–193
8. Düntsch I, Gediga G (2000) Rough set data analysis–a road to non-invasive knowledge discovery
9. Liu H, Yu L (2005) Toward integrating feature selection algorithms for classification and clustering. Knowl Data Eng IEEE Trans 17(4):491–502
10. Zhang F, Chan PPK, Biggio B, Yeung DS, Roli F (2015) Adversarial feature selection against evasion attacks
11. Pitt E, Nayak R (2007) The use of various data mining and feature selection methods in the analysis of a population survey dataset. In: Proceedings of the 2nd international workshop on Integrating artificial intelligence and data mining, vol 84, pp 83–93
12. Wang A, An N, Chen G, Li L, Alterovitz G (2015) Accelerating wrapper-based feature selection with K-nearest-neighbor. Knowl-Based Syst 83:81–91
13. Rais HM, Othman ZA, Hamdan AR (2007) Improved dynamic ant colony system (DACS) on symmetric traveling salesman problem (TSP). Int Conf Intell Adv Syst ICIAS 2007, pp 43–48
14. George A (2012) Anomaly detection based on machine learning: dimensionality reduction using PCA and classification using SVM. Int J Comput Appl Vol
15. Tsang C-H, Kwong S (2005) Multi-agent intrusion detection system in industrial network using ant colony clustering approach and unsupervised feature extraction. In: IEEE international conference on industrial technology ICIT 2005, pp 51–56
16. Gao H, Yang H, Wang X (2005) Ant colony optimization based network intrusion feature selection and detection, pp 18–21
17. Abd-Alsabour N, Randall M (2010) Feature selection for classification using an ant colony system. In: 2010 Sixth IEEE international conference on e-Science work, pp 86–91
18. Hiba Basim Alwan KRK-M (2013) Mixed variable ant colony optimization technique for feature subset selection and model selection, no 025, pp 24–31
19. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. ACM SIGKDD Explor Newsl 11:10–18
20. Tavallaee M, Bagheri E, Lu W, Ghorbani A-A (2009) A detailed analysis of the KDD CUP 99 data set. In: Proceedings of the second IEEE symposium on computational intelligence for security and defence applications 2009