AUT Journal of Modeling and Simulation

AUT J. Model. Simul., 51(1) (2019) 33-44
DOI: 10.22060/miscj.2019.14535.5110

Intrusion detection system using an ant colony gene selection method based on information gain ratio using fuzzy rough sets

Mohammad Masoud Javidi[1*], Sedighe Mansouri[2]

[1] Associate professor, Faculty of Mathematics and Computer Sciences, Shahid Bahonar University of Kerman, Kerman, Iran
[2] Master of science, Faculty of Mathematics and Computer Sciences, Shahid Bahonar University of Kerman, Kerman, Iran

ABSTRACT: With the development of network-based technologies, intrusion detection plays an important role in modern computer systems. Intrusion Detection System (IDS) is used to achieve higher security, and detect abnormal activities in computers or networks. The efficiency of intrusion detection systems mainly depends on the dimensions of data features. So, in the implementation of the IDS, by applying the feature selection phase irrelevant and redundant features are eliminated, and as a result, the speed and accuracy of the intrusion detection system increases. Applying appropriate search strategy and evaluation measure are significantly effective to feature selection. In this paper, we propose a feature selection method which uses a combination of filter and wrapper feature selection method. This method applies a modified ant colony algorithm as a search strategy on filter phase and fuzzy rough sets to calculate the information gain ratio and acquire the evaluation measure in the ant colony algorithm. Then, on the wrapper phase the minimal subsets of features with first order and second order accuracies are selected. To confirm the efficiency of our proposed method, we compared this method with three other methods and with a method which is based on artificial neural networks. Finally, we compared the proposed method with an ant colony optimization based method. Considering the results, the proposed method, on average, has a higher accuracy than the other methods and also selects a subset of features which have a minimum length.

Review History:
Received: 7 June 2018
Revised: 19 Junuary 2019
Accepted: 14 May 2019
Available Online: 1 June 2019

Keywords:
IDS
feature selection method
fuzzy rough sets
ant colony optimization

1. Introduction

A computer network includes the security of computer network infrastructure. Network security would be implemented with regards to network administrator policies. These policies are designed to protect the network resources against the unauthorized accesses [1]. The intrusion detection system (IDS) is used to secure computers and networks. Intrusion detection system plays an important role in detecting attacks and searching known or malicious activities in network traffic, and alert whenever a suspicious activity is detected [2]. There are two approaches for implementing intrusion detection: first, misuse detection or signature-based detection, and second, anomaly detection. In misuse detection, IDS compares current behavior of the network to a large database of attack signatures, then the system will alert once a match is detected. This approach can identify all known attacks with low false positive rate. However, when signatures are unknown, or the attack differs from the signature pattern the misuse detection would not be an appropriate method. In anomaly detection, the system administrator defines a baseline profile for normal network behavior such as traffic load, standard packet size, etc., and if the current behavior disagrees with this profile, it is detected as an intrusion [3]. IDS deals with huge amount of data containing both irrelevant and redundant features which might decrease the speed and detection rate of the system. Feature selection is used to overcome this problem and increase the efficiency. By feature selection irrelevant and redundant features would be removed, and an optimal subset of features would be selected. This results in improvement in processing time, classification performance and prediction accuracy [4].

The feature selection methods could be classified into filter, wrapper, embedded and hybrid methods. In filter method, goodness of the genes will be evaluated based on their intrinsic characteristics without applying any learning models. In wrapper method, learning algorithm is considered as a part of mentioned method. Here, the classification accuracy is used to select the features as evaluation criteria. The wrapper approach is more precise than filter approach and has more computation [5]. In embedded method, the learning is not separated from the feature selection part; so that the structure of the class of considered functions plays a crucial role [6]. The hybrid method combines filter and wrapper techniques [5].

The rest of this paper is organized as follows. In section 2, related works are discussed. In Section 3, some basic notations of rough set theory and information measures in rough and fuzzy-rough set theories are reviewed. The

*Corresponding author's email: javidi@uk.ac.ir

proposed method is presented in Section 4. In Section 5, the results of experiments are explained. Finally, Section 6 concludes the paper.

1-1- Related work

Employing the feature selection process to select an optimal subset of features instead of using the entire feature space is being widespread in many researches on the field of intrusion detection in computer networks [7]. Chung and Wahid [8] presented a hybrid intrusion detection system which applies Intelligent Dynamic Swarm based Rough Set (IDS-RS) and Simplified Swarm Optimization (SSO) which is a new version of Particle Swarm Optimization (PSO) and includes a new Weighted Local Search (WLS) strategy. IDS-RS method selects the most relevant features using a weighted sum fitness function. This work has selected six features out of 41 features containing in KDD cup 99 dataset. The acquired classification accuracy is 93.3%, which is not good enough. De la Hoz et al. [9] used a feature selection approach based on the NSGA-II4 as a feature search strategy and the Growing Hierarchical Self-Organizing Maps (GHSOM) as a classifier. They applied a fitness function based on the Jaccards coefficient which is a similarity measurement. The experiments are performed on the NSL-KDD datasets. This method selected 25 relevant features by 99.12% of classification accuracy which means the method is highly accurate, but the length of the selected subset is not satisfactory. Kang and Kim [10] proposed a wrapper method for feature selection based on a Local Search Algorithm (LSA) and the k-means clustering algorithm. This method has used the accuracy of k-means clustering as a cost function to measure the goodness of the feature subset generated by LSA. In order to avoid over fitting, Kang and Kim used MLP to evaluate the performance of the selected subset of features. The experiments are performed over the NSL-KDD datasets. The accuracy of classification and DR by means of this method are high but FAR is low. The methods mentioned so far have acceptable performances in solving the problem of intrusion detection. However, they still are not able to create a system that detects all attacks without any false alarms. In this paper, we have proposed a hybrid feature selection method which applies ant colony and fuzzy rough sets to calculate information gain ratio as evaluation criterion. Therefore, it has high accuracy and minimum subset length. After selecting an optimal subset of features, this subset is used in IDS. Applying dependency degree might be useful in selecting a subset of features as evaluation criteria, and also maintains the meaning of the features and rarely depends on other features; but it would not be appropriate in real life applications where the purpose is to acquire high classification accuracy [11]. Moreover, there is a tendency in gain criteria to select the feature with more refined partition. This fact encourages us to offer gain ratio as an improved version of gain based on fuzzy rough sets. Compared to the other methods, the proposed method has high classification accuracy and also chooses a subset with a minimum length. Additionally, there are some long loops in the implementation of the method which may result in high time complexity.

2. Theoretical or experimental modeling

2-1- Some basic notations

In this section, we have briefly described the theory of rough set and information measures in rough and fuzzy-rough sets theory. Rough set theory is proposed by Pawlak [12]. The basis of rough set theory is the concept of crisp equivalence class. A crisp equivalence class contains samples from different output classes. Moreover, the different elements in an equivalent class may have different degrees of belongingness to the output classes. To make a decision facing situations in which vagueness and indiscernibility are present, a combination of fuzzy and rough sets could be useful. A fuzzy similarity relation replaces an equivalence relation in rough sets to create fuzzy-rough sets. The detailed description of this process is presented in the following. This theory has been considered from the beginning, and has been used in various fields of data analysis such as banking [13], economics and finance [14], medical imaging [15], medical diagnosis [16], and data mining [17].

2-1-1- Basic rough set notation

Let, $IS =< U, A, V, f >$, be an information system, where U is a nonempty set of finite object, A is a finite set of attributes or genes, and V is the union of attribute domains, where V_a is the set of values for the attributes a; $f : A \times U \to V$ is an information function that appropriate special values from the domains of attribute to object. If $P \subseteq A$, then an associated indiscernibility equivalence relation, IND (P), is defined as [18]:

$$IND(P) = \{(x, y) \in U^2 \mid \forall a \in P f(a, x) = f(a, y)\} \quad (۱)$$

Since IND(P) is a reflexive, symmetric, and transitive relation, it is an equivalence relation. Therefore, IND(P) can create a partition on U that could be denoted by $U / IND(P)$ or more simply U / P and represents an equivalence class of IND(P) containing x. The lower and upper estimates for $X \subset U$, respectively are defined as follows [18]:

$$P \downarrow X = \{x \in U \mid [x]_P \subseteq X\} \quad (۲)$$

$$P \uparrow X = \{x \in U \mid [x]_P \cap X \neq \varnothing\} \quad (۳)$$

Based on the lower and upper estimates, the boundary regain is defined as follows [18]:

$$BND_P(X) = P \uparrow X - P \downarrow X \quad (۴)$$

2-1-2- Information measures in rough set theory

Assume Xi ⊆U / IND (P) and X j ⊆U / IND (Q) are partitions of U which are respectively induced by P and Q. The probability distribution of Xi is defined as follows. The probability distribution of XiX j is defined as Eq. (6), where |.| denotes the cardinality [18].

$$P(X_i)=\frac{|X_i|}{|U|} \quad (۵)$$

$$P(X_iX_j)=\frac{|X_i\cap X_j|}{|U|} \quad (۶)$$

Definition 1: If $IS=<U,A,V,f>$ is an information system, B is a subset of A and $X_i\in U/B$, then the Shannon's entropy H (B) of B is defined as [18]:

$$H(B)=-\sum_{i=1}^{n}P(X_i)\log P(X_i)=-\sum_{i=1}^{n}\frac{|X_i|}{|U|}\log\frac{|X_i|}{|U|} \quad (۷)$$

Definition 2: In information system $IS=<U,A,V,f>$, the join entropy of P and Q is defined as [18]:

$$H(PQ)=H(P\cup Q)=-\sum_{i=1}^{n}\sum_{j=1}^{m}P(X_iX_j)\log P(X_iX_j)$$
$$=-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|X_i\cap X_j|}{|U|}\log\frac{|X_i\cap X_j|}{|U|} \quad (۸)$$

That in relation $X_i\in U/P$, $X_j\in U/Q$ and $P,Q\subseteq A$.

Definition 3: the conditional entropy of D on condition B for decision system $DS=<U,C\cup D,V,f>$ is defined as [18]:

$$(D|B)=-\sum_{i=1}^{n}P(X_i)\sum_{j=1}^{m}P(X_j|X_i)\log P(X_j|X_i)$$
$$=-\sum_{i=1}^{n}\frac{|X_i|}{|U|}\sum_{j=1}^{m}\frac{|X_i\cap X_j|}{|U|}\log\frac{|X_i\cap X_j|}{|U|} \quad (۹)$$
$$=-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|X_i\cap X_j|}{|U|}\log\frac{|X_i\cap X_j|}{|U|}$$

That in this relation, B is a subset of C, and C is the condition attribute set; $X_i\in U/B$ and $X_j\in U/D$, where D is the decision attribute.

Definition 4: The mutual information of B and D is defined as follows [18]:

$$I(B;D)=H(D)-H(D|B) \quad (۱۰)$$

Definition 5: The gain of attribute $a\in C-B$ is defined as [18]:

$$Gain(a,B,D)=I(B\cup\{a\};D)-I(B;D)$$
$$=H(D|B)-H(D|B\cup\{a\}) \quad (۱۱)$$

Definition 6: The mutual information gain ratio of attribute a, is defined as [18]:

$$Gain\_Ratio(a,B,D)=\frac{Gain(a,B,D)}{H(\{a\})}$$
$$=\frac{I(B\cup\{a\};D)-I(B;D)}{H(\{a\})} \quad (۱۲)$$

2-1-3- Information measures in fuzzy-rough set theory

In fuzzy rough sets it is essential to define a fuzzy equivalence relation. If $\tilde{R}$ satisfies the following conditions, it would be considered as a fuzzy equivalence relation.

Reflectivity: $\tilde{R}(x,y)=1,\forall x\in X$

Symmetry: $\tilde{R}(x,y)=\tilde{R}(y,x),\forall x,y\in X$

Transitivity: $\tilde{R}(x,y)\geq\min\{\tilde{R}(x,y),\tilde{R}(y,z)\}$

The $M(\tilde{R})$ represents a relation matrix for $x_i,x_j\in X$, that $\tilde{R}$ is a fuzzy equivalence relation defined on a nonempty finite set X.

$$M(\tilde{R})=\begin{pmatrix}r_{11}&\cdots&r_{1n}\\ \vdots&\ddots&\vdots\\ r_{n1}&\cdots&r_{nn}\end{pmatrix} \quad (۱۳)$$

Here, $r_{ij}\in[0,1]$ is the relation value of x_i and x_j which can be written as . Considering the crisp rough set model, if x_i equals to x_j with respect to the crisp equivalence relation R then $r_{ij}=1$; otherwise, $r_{ij}=0$. A similarity function that has been used to calculate the equivalence relation is shown by Eq. (14), where x_i and x_j are attribute values of two objects on attribute a ; $a_{\max}$ and $a_{\min}$ are maximal and minimal values of attribute a respectively [18].

$$r_{ij}=\begin{cases}1-4\times\dfrac{|x_i-x_j|}{|a_{\max}-a_{\min}|}, & \dfrac{|x_i-x_j|}{|a_{\max}-a_{\min}|}\leq 0.25\\ 0 & \end{cases} \quad (۱۴)$$

Two important operators in the fuzzy equivalence relation which are also useful for implementing fuzzy theory are defined by [18]:

$$\tilde{R}=\tilde{R}_1\cup\tilde{R}_2\Leftrightarrow\tilde{R}(x,y)=\max\left\{\tilde{R}_1(x,y),\tilde{R}_2(x,y)\right\}$$

$$\tilde{R}=\tilde{R}_1\cap\tilde{R}_2\Leftrightarrow\tilde{R}(x,y)=\min\left\{\tilde{R}_1(x,y),\tilde{R}_2(x,y)\right\}$$

Definition 7: The fuzzy partition of the universe U generated by $\tilde{R}$, is defined as [18]:

$$U/\tilde{R}=\left\{[x_i]_{\tilde{R}}\right\}_{i=1}^{n} \quad (۱۵)$$

Here, $\tilde{R}$ is a fuzzy equivalence relation and $[x]_{\tilde{R}}$ is the fuzzy equivalence class equal to $\frac{r_{i1}}{x_1}+\frac{r_{i2}}{x_2}+\ldots+\frac{r_{in}}{x_n}$.

Definition 8: The cardinality is defined as [18]:

$$|[x_i]_{\tilde{R}}|=\sum_{j=1}^{n}r_{ij} \quad (۱۶)$$

Definition 9: Information quantity of the fuzzy attribute set or the fuzzy equivalence relation is defined as [18]:

$$H(\tilde{R}) = -\frac{1}{n}\sum_{i=1}^{n}\log\frac{|[x_i]_{\tilde{R}}|}{n} \quad (۱۷)$$

Definition 10: The joint entropy of B and E is defined as [18]:

$$H(BE) = H(\tilde{R}_B\tilde{R}_E) = -\frac{1}{n}\sum_{i=1}^{n}\log\frac{|[x_i]_{\tilde{B}} \cap [x_i]_{\tilde{E}}|}{n} \quad (۱۸)$$

Where FIS < U, A, V, f > is a fuzzy information system; A is the attribute set; B and E are two subsets of A.

Definition 11: While FIS <U, A, V, f > is a fuzzy decision system, C is the condition attribute set, D is the decision attribute and B □C. The condition entropy D on condition B could be calculated as follows [18]:

$$\tilde{H}(D \mid B) = -\frac{1}{n}\sum_{i=1}^{n}\log\frac{|[x_i]_{\tilde{B}} \cap [x_i]_{\tilde{D}}|}{|[x_i]_{\tilde{B}}|} \quad (۱۹)$$

In above-mentioned relation, $[x_i]_B$ and $[x_i]_D$ are fuzzy equivalence classes containing xi generated by B and D, respectively.

Definition 12: The mutual information of B and D is defined as [18]:

$$\tilde{I}(B;D) = \tilde{H}(D) + \tilde{H}(B) - \tilde{H}(BD) \quad (۲۰)$$

Definition 13: In decision system FDS <U,C□D,V, f > , $\forall a \in C - B$ the gain of attribute a, can be defined as [18]:

$$Gain(a,B,D) = \tilde{I}(B \cup \{a\};D) - \tilde{I}(B;D) \quad (۲۱)$$

Definition 14: According to the definition 13, the mutual information gain ratio of attribute a, can be defined as [18]:

$$GainRatio(a,B,D) = \frac{Gain(a,B,D)}{\tilde{H}(\{a\})} = \frac{\tilde{I}(B \cup \{a\};D) - \tilde{I}(B;D)}{\tilde{H}(\{a\})} \quad (۲۲)$$

2-1-4 Data set

The data set which is used here, is KDD Cup99. It contains five million training data records and two million testing data records. Each record has 41 features. This data set consist of 4 kinds of different attacks; DOS, R2L, U2R and PRB. In table 1, the list of KDD cup 99 features is presented.

2-1-5- Attacks in KDD Cup99

There are 4 kinds of different attacks in KDD Cup99:

•Denial of Service Attacks (DOS)

A DOS attack is a type of attack in which the hacker makes a computing or memory resources too busy or too full to serve legitimate networking requests, and as a result denies users' access to a machine [20].

•Remote to User Attacks (R2L)

A remote to user attack is an attack in which a user sends packets to a machine over the internet with no access to it. This is done by the purpose of exposing the machines vulnerabilities and exploiting privileges which a local user would have on his/her computer [21].

•User to Root Attacks (U2R)

These attacks are exploitations in which the hacker starts off on the system with a normal user account and attempts to abuse vulnerabilities in the system to gain super user privileges [21].

•Probing (PRB)

Probing is an attack in which the hacker scans a machine or a networking device to determine weaknesses or vulnerabilities which could be exploited later to compromise the system [21].

2-2- Proposed method

In this section, a new filter-wrapper approach for feature selection in fuzzy-rough sets is described. In this approach, the filter phase employs a modified ACO search strategy which is able to do feature selection as a multi-modal problem. The wrapper phase includes a learning model which evaluates the chosen subsets of features from the filter phase and selects the best subset, then calculates pheromones changes in selected subsets. Choosing the subsets of features with first and second maximum accuracies as candidate subsets for minimal data reductions is a contribution of this work; so each chosen minimal subset would have short length along with acceptable accuracy value. Consequently, the approach would satisfy both increasing the accuracy and decreasing the length of reduced subsets, concurrently.

In detail, to implement the approach, we need the feature selection problem space to be considered in the form of a complete non-directed graph. The nodes, indicating the features and edges, represent the probability of choosing the next node. The algorithm starts with the production of k number of ants, which is half the number of features. The following steps are done to complete each ant's tour:

1- Initialize ants with random and different nodes.

2- For each ant k, consideration is that the set S_K includes all the nodes without initial node, as accessible locations.

3- The ant k chooses the next node according to the transition rule. This rule has been further described on the next section.

4- The selected node is removed from S_K.

5- For each ant k, the third and fourth stages are repeated until S_K is empty.

Table 1[19]. List of KDD Cup 99 features

Category	Label/feature name	Type	Description
1	1. Duration	Continuous	Length (number of seconds) of the connection
	2. Protocol-type	Discrete	Type of the protocol, e.g., tcp, udp, etc.
	3. Service	Discrete	Network service on the destination, e.g., http, telnet, etc.
		Discrete	Normal or error status of the connection
	4. Flag	Continuous	Number of data bytes from source to destination
	5. Src-bytes	Continuous	Number of data bytes from destination to source
	6. Dst-bytes	Discrete	1 If connection is from/to the same host/port; 0 otherwise
	7. Land		Number of ‘‘wrong’’ fragments
		Continuous	Number of urgent packets
	8. Wrong-fragment	Continuous	
	9. Urgen		
2	10. Hot	Continuous	Number of ‘‘hot’’ indicators (hot: number of directory accesses, create and execute program)
	11. Num-failed-logins	Continuous	Number of failed login attempts
	12. Logged-in	Discrete	1 If successfully logged-in; 0 otherwise
	13. Num-compromised	Continuous	Number of ‘‘compromised’’ conditions (compromised condition: number of file/path not found errors and jumping commands)
	14. Root-shell	Discrete	1 If root-shell is obtained; 0 otherwise
	15. Su-attempted	Discrete	1 If ‘‘su root’’ command attempted; 0 otherwise
	16. Num-root	Continuous	Number of ‘‘root’’ accesses
	17. Num-file-creations	Continuous	Number of file creation operations
	18. Num-shells	Continuous	Number of shell prompts
	19. Num-access-files	Continuous	Number of operations on access control files
	20. Num-outbound-cmds	Continuous	Number of outbound commands in an ftp session
	21. Is-host-login	Discrete	1 If the login belongs to the ‘‘hot’’ list; 0 otherwise
	22. Is-guest-login	Discrete	1 If the login is a ‘‘guest’’login; 0 otherwise
3	23. Count	Continuous	Number of connections to the same host as the current connection in the past 2 s
	24. Srv-count	Continuous	Number of connections to the same service as the current connection in the past 2 s (same-host connections)
	25. Serror-rate	Continuous	% Of connections that have ‘‘SYN’’ errors (same-host connections)
	26. Srv-serror-rate	Continuous	% Of connections that have ‘‘SYN’’ errors (same-service connections)
	27. Rerror-rate	Continuous	% Of connections that have ‘‘REJ’’ errors (same-host connections)
	28. Srv-rerror-rate	Continuous	% Of connections that have ‘‘REJ’’ errors (same-service connections)
	29. Same-srv-rate	Continuous	% Of connections to the same service (same-host connections)
	30. Diff-srv-rate	Continuous	% Of connections to different services (same-host connections)
	31. Srv-diff-host-rate	Continuous	% Of connections to different hosts (same-service connections)
4	32. Dst-host-count	Continuous	Count for destination host
	33. Dst-host-srv-count	Continuous	Srv_count for destination host
	34. Dst-host-same-srv-rate	Continuous	Same_srv_rate for destination host
	35. Dst-host-diff-srv-rate	Continuous	Diff_srv_rate for destination host
	36. Dst-host-same-src-port-rate	Continuous	Same_src_port_rate for destination host
	37. Dst-host-srv-diff-host-rate	Continuous	Diff_host_rate for destination host
	38. Dst-host-serror-rate	Continuous	Serror_rate for destination host
	39. Dst-host-srv-serror-rate	Continuous	Srv_serror_rate for destination host
	40. Dst-host-rerror-rate	Continuous	Rerror_rate for destination host
	41. Dst-host-srv-rerror-rate	Continuous	Srv_serror_rate for destination host

6- The best answer ever acquired has been saved.

After each ant completes its tour, the pheromone would be updated on the routes traversed from the origin to the destination according to the algorithm explained in section 4-2. At the end of each iteration, the best observed solutions thus far are kept; i.e., in each iteration, we consider the subsets of the features which have maximum accuracies as the best candidate subsets. We preserve the subsets which have the first and the second maximum accuracies among the best candidate subsets from the first iteration to the current. Then, we consider the minimal subsets within the preserved subsets as the bests in all iterations. Since the wrapper method utilizes a learning model, feature selection based on wrappers

increments the accuracy of the model; however, this method increases order of mathematical complexity. In this method, instead of evaluating the features separately, the subsets found by filter are evaluated by wrapper model to decrease the complexity. The output of the wrapper model (the accuracy of the classifier) would be a criterion for goodness evaluation of found subsets. After the end of each run, the best acquired solution from the first iteration will be saved as an optimal solution. In addition to the detection of high quality subsets of features, finding more than a single solution in just one run is one of the advantages of this method compared to other ones.

2-2-1- Transition rule and feature deletion

The transition rule introduced in [22] is used for exploring the nodes' space. Node j, as a candidate for selection, is selected with probability equal to 0.5, using the following relation:

$$p_{ij}^{k}=\begin{cases}1, & j=\arg\max\left\{\tau_{ij}^{\alpha}\eta_{ij}^{\beta}\right\}\\ 0, & otherwise\end{cases} \quad (۲۳)$$

If an ant selects a new node, that node is removed from the set of available nodes. Additionally, if the candidate node j is not selected, the candidate node will also be removed from the set of available nodes. In this case, the following relation in the roulette wheel mechanism, as the probability of selecting available nodes, is used to select the next node.

$$p_{ij}^{k}(t)=\begin{cases}\dfrac{[\tau_{ij}]^{\alpha}.[\eta_{j}]^{\beta}}{\sum_{x\in S_K}[\tau_{ix}]^{\alpha}.[\eta_{x}]^{\beta}}, & j\in s_k\\ 0, & otherwise\end{cases} \quad (۲۴)$$

In both of the above-mentioned equations, α = 0.5 and β = 1, and the initial value of τ_{ij} is equal to 0.1. By selecting each node in the roulette wheel mechanism, the selected node and all nodes prior to it, In this method, $\eta_j=\widetilde{GainRatio}(j,N_K,D)$ is calculated by Eq. (22) as heuristic information. N_K is regarded as a set of selected nodes by ant k, and is the pheromone value of edge ij.

2-2-2- Pheromones updating Rules

After each individual ant created its own complete tour, the pheromone is updated on the travelled path from the beginning to the end, as follows:

1- On each edge of the complete graph, the pheromone evaporates according to the equation (25).

2- In each iteration, the pheromone on the path is updated according to (26) and (27).

3- In order to maintain the best answers ever acquired, the pheromone on the best path within all repetitions would be updated according to (28).

$$\tau^{new}=(1-\rho).\tau^{old} \quad (۲۵)$$

$$\Delta\tau_{ij}=\frac{\gamma'_{N_k}}{lenght(N_k)} \quad (۲۶)$$

$$\tau_{ij}^{new}=\begin{cases}\tau_{ij}^{old}+\Delta\tau_{ij}, & if \quad ij\in BF\\ \tau_{ij}^{old}+\varphi*\Delta\tau_{ij}, & otherwise\end{cases} \quad (۲۷)$$

$$\tau_{ij}^{new}=\tau_{ij}^{old}+\varphi*\gamma'_{N_k} \quad (۲۸)$$

Where □=0.5, □=0.2 and BF is the best path traversed in the current iteration. γ'_{N_k} is the accuracy of classifier as the output of the learning model. Figure 1, represents the graph which is created in the proposed feature selection method.

3. The results and discussion

To implement the proposed method, we utilized the R statistical software on a Windows platform having configuration Intel core 5 Duo CPU 2.49 GHZ, 2 GB RAM. The results of these experiments have been compared with three other good methods including MMIFS [19], LCFS and FFSA [19]. MMIFS and LCFS are two dependency based feature selection algorithms with different evaluation functions. LCFS method is based on the linear correlation coefficient, while MMIFS method is based on the mutual information. Linear correlation is used in detecting features having near linear correlation to the system output. However, in the real world, since the correlations are not always linear, this method could not be suitable. While, the MMIFS method can measure arbitrary relations between features [19], FFSA is a mutual information-based forward feature selection method [19]. The results are expressed in terms of time, accuracy and

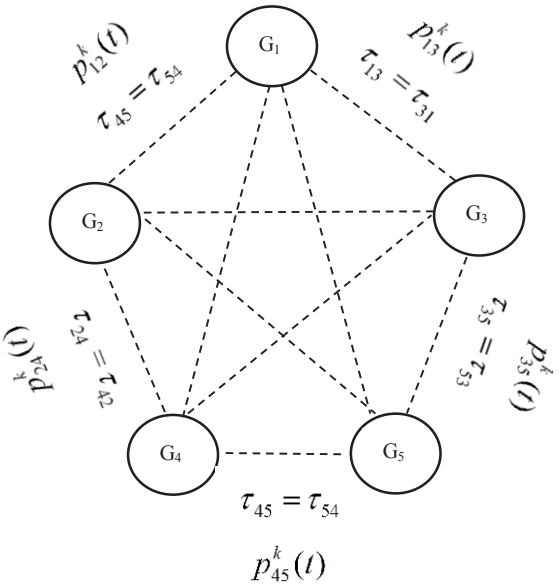


Fig. 1. The graph representation for the feature selection problem. $p_{ij}^{k}(t)$ is the probability of selecting the available nodes. And τ_{ij} denotes the pheromone value of edge ij.

a criterion named ψ_B. This criterion is defined as follows, and it has been examined in the results as well.

$$\psi_B = \frac{accuracy(B)}{length(B)}$$

Where B is a subset of features. By increasing the classification accuracy and reducing the length of the selected subset, ψ_B increases. This indicates an increase in the efficiency of the method, since the efficiency of the feature selection is based on accuracy and number of selected features.

In figure 2, the computational time of the proposed method has been compared to the other methods. According to the figure 2, the consuming time of the proposed method is longer than the other algorithms. The high execution time of this algorithm is due to the fact that the implementation of the ant colony algorithm requires several nesting loops with a high repetition rate. Moreover, the number of high repetitions is due to the fact that network data sets such as KDD cup 99, usually have a high number of features. Another reason for the high execution time of this algorithm which could be indicated is that high-dimensional square matrices are needed to calculate entropy and information gain ratio; the calculations of these matrices are time consuming.

The subset obtained from the feature selection phase in the proposed method is used as the input to a PLSSVM intrusion detection system, and its performance such as its accuracy is calculated. PLSSVM (Partial Least Squares Support Vector Machine) is a SVM classifier and performs on the score of the partial least squares. After introducing the input to the mentioned system and calculating its accuracy, the PLSSVM accuracy is compared to the accuracies of some other SVM-based classifiers. In table 2, detection rate, false positive rate, accuracy, and the number of detection errors for four methods have been compared under various intrusion classes. According to the table 2, among FFSA, MMIFS and LCFS methods, the average of detection rate for MMIFS is the highest. Therefore the average of detection rate for the proposed method has been increased by approximately 0.8%

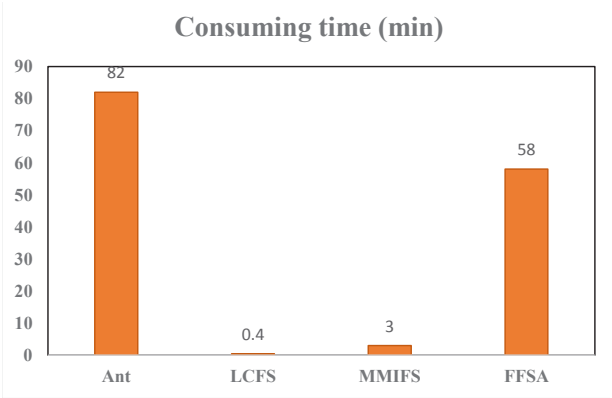


Fig. 2. Consuming time (in minute) of different feature selection algorithms

compared to MMIFS method. Also, among FFSA, MMIFS and LCFS methods, the average of false detection rate for FFSA is the lowest. As the result, the average of false detection rate of the proposed method has been decreased by almost 17.1% compared to FFSA algorithm. Furthermore, in mentioned methods, the accuracy of the classifier of the FFSA algorithm is the highest; therefore, the average of accuracy of the proposed method has been increased by 0.6% compared to FFSA method. The improvement of the proposed method is due to the use of an ant colony algorithm and information gain rate by fuzzy rough sets which has been used as an evaluation measure.

In Figures 3-7, the criteria of four methods are shown in normal situation and different types of attacks. In evaluation of the efficiency of feature selection methods, along side the classifier accuracy criteria, the length of the selected subset is also important. Therefore, the criteria would be more appropriate. The high value of this criterion for a feature selection method represents the high efficiency of the method. According to these figures, the criteria of the proposed method is higher than the other methods, which is increased in average by 16.5% compared to MMIFS method. The criteria is also increased in average by 81% compared to FFSA algorithm. Therefore, the proposed method has high efficiency.

The proposed method in wrapper phase, selects two subsets with high accuracies and among them choses a subset with the minimum length. So, this algorithm selects a subset with high accuracy and minimum length which leads to high criteria and efficiency.

In table 3, the accuracy of the PLSSVM is compared to some SVM-based methods on feature selection such as SVM, Bayesian and FNT. The results show that all of them are effective classification methods. An LSSVM is a regularized reformulation to the standard SVM. PLSSVM is effective in avoiding local minima in SVM problems [19]. According to this table, PLSSVM model seems to be more promising. PLSSVM outperforms the SVM, Bayesian and FNT in detecting Normal, Probe and R2L classes. In detecting the DOS and U2R classes, all four classifiers work well and there are no significant differences.

The results show that this method has high accuracy and minimum subset length. This is due to the use of ant colony and fuzzy rough sets to calculate information gain ratio as evaluation criterion. However, this method has high computational time.

In table 4, the proposed method has been compared with an artificial neural network based method. In this approach the malicious network traffic is detected by artificial neural networks [23]. According to the table 4, the accuracy and recall of the proposed method is higher and its precision criteria is lower than the artificial neural network based method. Generally, the performance of the proposed method is almost equal to the performance of the artificial neural network based method.

In figure 8, a comparison between the proposed method and an ant colony-based method on detection rate is presented

Table 2. Performance of classification for attack and Normal classes in evaluation data (method: feature selection algorithm, DR: detection Rate, FPR: false positive rate, DE: the number of detection errors)

class	Method	DR (%)	FPR (%)	Accuracy (%)	DE(# records)
Normal	Ant rough set	99.95 ± 0.100	0.06 ± 0.171	99.92 ± 0.142	5
	FFSA	99.843 ± 0.143	0.25 ± 0.22	99.80 ± 0.1	13
	MMIFS	99.92 ± 0.062	0.08 ± 0.062	99.89 ± 0.035	7
	LCFS	99.73 ± 0.257	2.24 ± 2.96	99.37 ± 0.66	51
	All features	99.743 ± 0.452	10.48 ± 9.72	96.81 ± 2.19	210
Dos	Ant rough set	89.10 ± 0.03	0.02 ± 0.1	98.86 ± 0.03	77
	FFSA	90.02 ± 0.2	0.02 ± 0.3	99.00 ± 0.25	70
	MMIFS	85.81 ± 0.09	0.03 ± 0.2	98.9 ± 0.15	91
	LCFS	87.84 ± 0.5	1.1 ± 1.5	98.83 ± 1.0	130
	All features	85.81 ± 1.0	10.0 ± 6.1	97.64 ± 2.1	230
Probe	Ant rough set	99.99 ± 0.01	0.17 ± 0.1	99.90 ± 0.023	11
	FFSA	92.9 ± 0.5	0.19 ± 0.21	99.09 ± 0.3	61
	MMIFS	99.97 ± 0.05	0.19 ± 0.12	99.83 ± 0.045	11
	LCFS	57.15 ± 1.0	0.2 ± 0.9	95.8 ± 1.1	310
	All features	78.58 ± 3.2	11.1 ± 0.5	95.1 ± 1.8	350
R2L	Ant rough set	99.98 ± 0.05	0.2 ± 0.08	99.95 ± 0.15	5
	FFSA	99.72 ± 0.09	0.2 ± 0.1	99.79 ± 0.19	14
	MMIFS	99.98 ± 0.08	0.3 ± 0.1	99.91 ± 0.07	6
	LCFS	93.57 ± 1.08	0.07 ± 0.8	99.61 ± 1.3	77
	All features	99.7 ± 0.29	62.01 ± 10.0	84.24 ± 9.02	1056
U2R	Ant rough set	96 ± 1.7	5.98 ± 0.1	95.26 ± 0.3	402
	FFSA	90 ± 3.6	7.1 ± 0.32	93.16 ± 0.5	458
	MMIFS	95 ± 1.01	9.66 ± 0.4	90.32 ± 0.5	648
	LCFS	50 ± 5.00	5.69 ± 1.21	94.20 ± 2.0	388
	All features	95 ± 6.2	5.46 ± 0.65	94.56 ± 0.1	364

[24]. Considering this figure, the proposed method has a higher detection rate in normal, probe, R2L and U2R classes. Therefore, it would be discernible that the proposed method is more effective in detecting such abnormal activities.

In table 5, the proposed method and the ant colony-based method has been compared to each other with regards to the accuracy, FPR, precision and recall measures. According to this table, the proposed method has a lower accuracy but has a higher FPR, precision and recall measures. Therefore, due to the detection rate, FPR, precision and recall measures, despite the lower accuracy of the proposed method, in general, this method has a higher efficiency than the ant colony based method.

4. Conclusion and future works

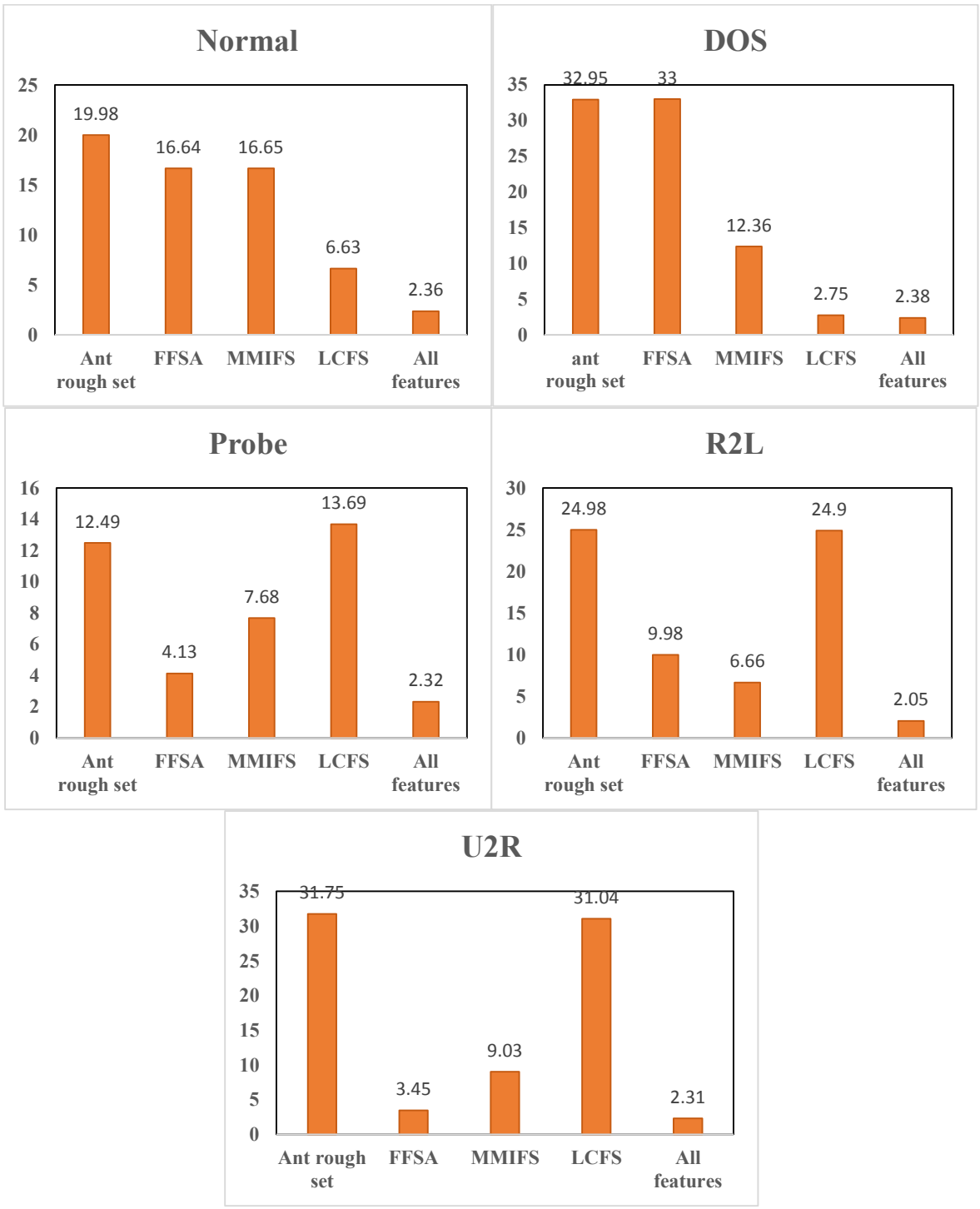
In this research, a feature selection method is proposed to eliminate redundant and additional features in KDD Cup 99 dataset. Then, by applying this deducted data set we modeled an intrusion detection system. In feature selection phase, a hybrid of filter and wrapper feature selection methods is introduced. In the filter section, features are obtained using the ant colony algorithm and information gain ratio which

Table 3. The evaluation results comparison with the other approaches according to accuracy rate (each row represent a method of intrusion detection)

method	Normal (%)	Doss (%)	Probe (%)	R2L (%)	U2R (%)
PLSSVM	99.92	98.86	99.90	99.95	95.26
SVM	99.45	98.70	99.69	99.88	95.99
Bayesian	99.58	98.53	99.55	99.21	76.59
FNT	99.30	98.44	99.37	99.10	95.85

Table 4. The evaluation results of the proposed method and the artificial neural network based method

Method	Accuracy (%)	Precision	Recall (sensitivity)
Ant rough set	98.76	0.96	0.97
Neural network	98	0.97	0.95



Figs. 3-7. The ψ_B criteria of four methods in normal situation and different types of attacks

is calculated by fuzzy rough sets. Moreover, in the wrapper section, the feature set obtained from the filter phase is evaluated. Finally, the best subset of features is collected. The feature set acquired from the feature selection stage is used as the input of the intrusion detection system. The proposed method is implemented using the statistical software R. The results show that the proposed method is highly effective due to the high accuracy and rate; so that the average accuracy of the proposed method is increased by 0.6% compared to the best of aforementioned methods. The average of has increased by 16.5% compared to MMIFS method, however, this method has low runtime. For the future works, we will further study this field using our proposed feature selection methods as a preprocessing step in other learning methods. Additionally, we will do a comparison between our proposed feature selection method with other ones.

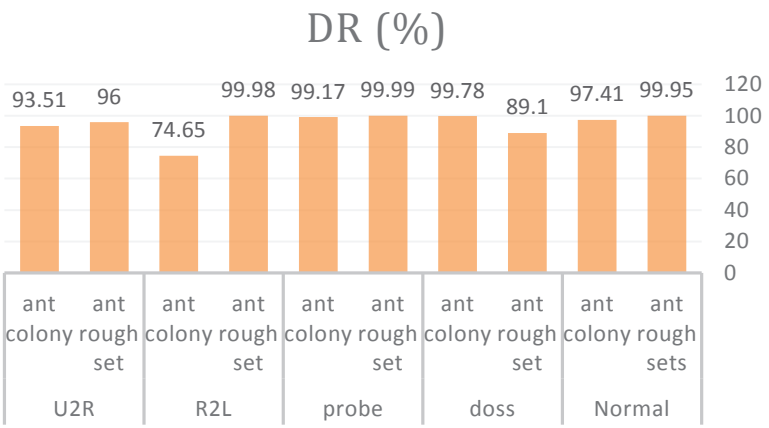


Fig. 8. A comparison based on the detection rate between the proposed method and ant colony based-method

Table 5. Comparison between the proposed method and ant colony based method based on accuracy, FPR, precision and recall measures

Method	Accuracy (%)	FPR	Precision	Recall
Ant rough set	97.76	1.286	0.96	0.97
Ant colony	98.9	2.59	0.52	0.92

References

[1] U. Ravale, N. Marathe, P. Padiya, "Feature Selection Based Hybrid Anomaly Intrusion Detection System Using K Means and RBF Kernel Function. ", In proceedings of the International Conference on Advanced Computing Technologies and Applications (ICACTA), pp. 428-435, 2015.

[2] A. A. Aburomman, M. B. I. Reaz, "A Novel Weighted Support Vector Machines Multiclass Classifier Based on Differential Evolution for Intrusion Detection Systems", Information Sciences, vol. 414, pp. 225-246, 2017.

[3] A. Sharma, I. Manzoor, N. Kumar, "A Feature Reduced Intrusion Detection System Using ANN Classifier", Expert Systems with Applications, vol. 88, pp. 249-257, 2017.

[4] Y. Zhu, J. Liang, J. Chen, Z. Ming, "An improved NSGA-III algorithm for feature selection used in intrusion detection", Knowledge-Based Systems, vol. 116, pp. 74-85, 2017.

[5] S. N. Ghazavi, T. W. Liao, "Medical data mining by fuzzy modeling with selected features", Artificial Intelligence in Medicine, vol. 43, pp. 195-206, 2008.

[6] T.N. Lal, O. Chapelle, J. Weston, A. Elisseeff, Embedded methods, in: I. Guyon, S. Gunn, M. Nikravesh, L.A. Zadeh (Eds.), Feature Extraction: Foundations and Applications. Studies in Fuzziness and Soft Computing, vol. 207, Springer, Berlin, Heidelberg, pp. 137–165, 2006.

[7] Ch. Khammassi, S. Krichen. "A GA-LR Wrapper Approach for Feature Selection in Network Intrusion Detection.", Computers & Security, 2017.

[8] Y. Y. Chung, N. Wahid, "A hybrid network intrusion detection system using simplified swarm optimization (sso)", Applied Soft Computing, vol. 12, pp. 3014–3022, 2012.

[9] E. De la Hoz, A. Ortiz, J. Ortega, A. Martínez-Álvarez, "Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps.", Knowledge-Based Systems, vol. 71, pp. 322–338, 2014.

[10] S. H. Kang, K. J. Kim, "A feature selection approach to find optimal feature subsets for the network intrusion detection system", Cluster Computing, pp. 1–9, 2016.

[11] P. Maji, P. Garai, "On fuzzy-rough attribute selection: Criteria of Max-Dependency, Max-Relevance, Min-Redundancy, and Max-Significance.", applied soft computing, vol. 13, pp. 3968-3980, 2013.

[12] Z. Pawlak, A. Skowron, "Rudiments of rough sets", Information sciences, vol.177, pp. 3-27, 2007.

[13] G.A. Montazer, S. ArabYarmohammadi, "Detection of phishing attacks in Iranian e-banking using a fuzzy–rough hybrid system", Applied Soft Computing, vol. 35, pp. 482–492, 2015.

[14] M. Podsiadło & H. Rybiński, "Rough sets in economy and finance", Transactions on Rough Sets XVII, pp. 109-173, 2014.

[15] C.H. Xie, Y.J. Liu, J.Y. Chang, "Medical image segmentation using rough set and local polynomial regression", Multimedia Tools and Applications, vol. 74, pp. 1885–1914, 2015.

[16] V. Prasad, T.S. Rao, M.S. Babu, "Thyroid disease diagnosis via hybrid architecture composing rough data sets theory and machine learning algorithms", Soft Computing, vol. 20, pp. 1179–1189, 2016.

[17] M.P. Francisco, J.V. Berna-Martinez, A.F. Oliva, M.A.A.

Ortega, "Algorithm for the detection of outliers based on the theory of rough sets", Decision Support Systems, vol. 75, pp. 63–75, 2015.

[18] J. Dai, Q. Xu, "Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification", Applied Soft Computing, vol. 13, pp. 1184-1199, 2012.

[19] F. Amiri, M. Rezaei, C. Lucus. A. Shakeri, N. Yazdani, "Mutual information-based feature selection for intrusion detection systems", Network and computer applications, vol. 34, pp. 1184-
1199, 2011.

[20] İ. Özçelik, R. R. Brooks, "Deceiving entropy based DoS detection", computers & security, vol. 48, pp. 234-245, 2015.

[21] Sh. Aljawarneh, M. Aldwairi, M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model", Computational Science, 2017.

[22] M. Dorigo, L. M. Gambardella, "A cooperative learning approach to the traveling salesman problem", IEEE Transactions on Evolutionary Computation, vol. 1, pp. 53-66, 1997.

[23] A. Shenfield, D. Day, A. Ayesh, "Intelligent intrusion detection systems using artificial neural networks.", the Korean Institute of Communications and Information Sciences, vol. 4, pp. 95-99, 2018.

[24] M. Hosseinzadeh, P. Kabiri. "Feature selection for intrusion detection system using ant colony optimization", International Journal of Network Security, vol. 18, pp. 420-432, 2016.

HOW TO CITE THIS ARTICLE

M. M. Javidi and S. Mansouri, Intrusion detection system using an ant colony gene selection method based on information gain ratio using fuzzy rough sets, AUT J. Model. Simul., 51(1) (2019): 33 - 44.

DOI: 10.22060/miscj.2019.14535.5110