Cross Layer Intrusion Detection System of Mobile Ad Hoc Networks using Feature Selection Approach

POONGOTHAI.T
Department of Information Technology
K.S.R College of Engineering, Tiruchengode 637 215
INDIA
poongothait@gmail.com

DURAISWAMY.K
Dean
K.S.Rangasamy College of Technology, Tiruchengode 637 215
INDIA

Abstract: - Existing intrusion detection system of Mobile ad hoc networks (MANET) examines all the features of network audit data. Some of the features of collected data may be redundant or irrelevant to the detection process. So it is vital to select the important features to increase the detection accuracy. This paper focuses on implementing two feature selection methods namely, Rough Set Theory (RST) and genetic algorithm (GA) combined with Support Vector Machines (SVM). Also the proposed system uses cross layer features instead of single layer features to maximize the performance. The results are validated through ns-2 simulations. The efficiency of the IDS is analyzed with varying network conditions by simulating routing attacks. The system has achieved an overall detection accuracy of detection with all features is 96.37, rough set feature selection is 97.34 and genetic feature selection is 98.22. Simulation results show that the proposed cross-layer approach aided by a combination of GA and SVM performs significantly better than other approach.

Key-Words: - Mobile Ad Hoc Networks, Intrusion Detection, Cross-Layer Design, Feature Selection, Rough Set Theory, Genetic Algorithm, Support Vector Machine.

1 Introduction

In recent years, Mobile Ad Hoc Networks (MANET) are one of the fastest growing areas of research and more popular technology in wireless network because of the increased usage of wireless devices. Unlike conventional networks, they do not have fixed infrastructure and centralized management. Each node in the network needs to act as a router as well as a host. Therefore, node cooperation is very much important for the network functioning. This creates a lot of security vulnerabilities called attacks in MANET. Attack prevention measures such as authentication and encryption are used to handle outside attacks, but they cannot detect inside attacks. The intrusion detection system acts as a second line of defense to detect inside intruders. Existing Intrusion detection technique developed for wired networks cannot be applied directly to MANET. The use of IDS developed for wired networks in order to safeguard the MANET is neither direct nor easy to perform. It needs to be restructured to suit the characteristics of MANET [1, 2, 3].

The main function of the IDS is to identify the intrusion from audit data collected from network.
The existing intrusion detection system of Mobile ad hoc networks (MANET) examines the activities or events of individual layer of network. But most of the attacks simultaneously exploit the vulnerabilities at multiple layer of the network. The features collected from single layer are not sufficient to detect the suspicious behavior [4, 5]. It is essential to collect the features from multiple layers to efficiently detect the abnormalities in the mobile ad hoc networks.

Also past work on intrusion detection system of MANET examines all the features of collected data. Some of the features may be redundant or contribute little to the detection process causing slow training and testing process. The redundant features increase the resource consumption of IDS and also affect the detection accuracy. The classifier cannot classify correctly because of the presence of irrelevant

features. So it is vital to select the important features to improve the performance of a classifier [6,7,8]. The contribution of the proposed system is

- To increase the detection rate and reduce the false alarm rate by collecting information from multiple layers.
- To evaluate the performance of cross-layer intrusion detection system by applying feature selection techniques.

Different techniques have been used for the purpose of feature selection. Literature shows that the performance of rough set theory [9, 10] and genetic algorithm [11, 12,13] is appreciable. Therefore, the proposed system uses two feature selection techniques namely, rough set theory and genetic algorithm. A nonlinear machine learning algorithm, Support Vector Machines is used as a classifier.SVM offers excellent detection accuracy compared to other machine learning approaches [14, 15, 16].

The rest of the paper is organised as follows. Section 2 describes feature selection problem. Section 3 explains the architecture of proposed system. Experimental results are described in section 4. Finally, Section 5 concludes the paper.

2 Feature Selection

Feature selection is the process selecting subset of features by eliminating irrelevant and uninformative features from original feature set. The advantage of the feature selection process is improving the performance of the learning algorithm and speed up the computation process of the resulting model [17]. The two main categories of feature selection are filter approach and wrapper approach [6, 18]. In filter approach, the features are selected before applying machine learning algorithm to the data set. In wrapper approach, the features are selected depending on the classifier. Filter approach is computationally efficient than wrapper approach. Wrapper approach yields better result because of the usage of machine learning algorithm for selecting optimal features.

A key problem faced by many classification algorithms is the selection of optimal features from the training data set. Feature selection significantly increases the performance of the classifiers. Intrusion Detection System is a typical classification problem. Feature selection plays a vital role in the development of building intrusion detection system. During this process, most effective features are extracted and the features containing false correlations or irrelevant features are eliminated.

The main goals of feature selection are

(1) Reducing the building and testing time of a classifier by minimizing the number of features.

(2) Maximizing the classification ability by removing irrelevant and redundant features.

(3) Increasing the data quality by which data understanding is improved.

(4) Minimizing storage requirements by reducing the dimensionality of the feature space.

2.1 Rough Set Theory

Rough set theory (RST) is an extension of classical set theory that supports approximations in decision making. This concept was introduced by Zdzislaw Pawlak [19] in the early 1980's. RST proposes a new mathematical approach to imperfect knowledge. It is an approximation of a vague concept by a pair of precise concepts, called lower and upper approximations, which are a classification of the domain of interest into disjoint categories. The lower approximation is a description of the domain objects which are known with certainty to belong to the subset of interest, whereas the upper approximation is a description of the objects which possibly belong to the subset. The approximations are constructed with regard to a particular subset of attributes or features.

RST is a mathematical tool for approximate reasoning for decision support. It can be used for feature selection and feature extraction. In rough set theory the data is represented as a table, called decision table. Rows of the decision table correspond to objects, and columns correspond to attributes. The class label is known as the decision attribute and the rest of the attributes known as the condition attributes.

The basic concept of the RST is the notion of approximation space, which is an ordered pair A= (U, R), where

U: nonempty set of objects, called universe

R: equivalence relation on U, called indiscernibility relation. If x, y $\in$ U and xRy then x and y are indistinguishable in A. Let X be a subset of U, i.e. $X \subseteq U$. $R(x)$ denote the equivalence class of R determined by element x.

The lower approximation of a set X with respect to R, and denoted by $R_*(X)$

$$R_* (X) = \{x: R(x) \subseteq X\}$$

The upper approximation of a set X with respect to R, and denoted by $R^* (X)$

$$R^* (X) = \{x: R(x) \cap X \neq \phi\}$$

The set of all objects which can be decisively classified neither as members of X nor as members of - X with respect to R is called the boundary region of a set X with respect to R, and denoted by $RN_R(X)$.

$$RN_R(X) = R^* (X) - R^* (X)$$

The lower approximation of a set is union of all granules which are entirely included in the set; the upper approximation is union of all granules which have non-empty intersection with the set; the boundary region of a set is the difference between the upper and the lower approximation of the set.

The following are the four basic classes of rough sets.

1. A set X is *roughly R-definable*, iff $R_* (X) \neq \phi$ and $R^* (X) \neq U$.
2. A set X is *internally R-undefinable*, iff $R_* (X) = \phi$ and $R^* (X) \neq U$.
3. A set X is *externally R-undefinable*, iff $R_* (X) \neq \phi$ and $R^* (X) = U$.
4. A set X is *totally R-undefinable*, iff $R_* (X) = \phi$ and $R^* (X) = U$.

RST classifies all the attributes into three categories: core attributes, reduct attributes and dispensable attributes. Core attributes have the essential information to make correct classification for the data set and should be retained in the data set; dispensable attributes are the redundant ones in the data set and should be eliminated; and reduct attributes are in the middle between. Depending on the combination of the attributes, in some cases, a reduct attribute is not necessary, while in other situations it is essential.RST can be used to combine the similar attributes and to reduce the number of attributes. So it can increase the processing speed and raises the detection rate.

2.2 Genetic Algorithm

Genetic Algorithm (GA) is a stochastic search method which is inspired by Darwinian natural selection and biological reproduction [20]. It maps the searching space in to genetic space. The searching space is encoded into a chromosome. The chromosome represents the problem to be solved. Each element of a chromosome represents a gene. All of the chromosomes make up a population and is represented by means of binary strings of 0's and 1's. The goodness of a chromosome is measured by using the fitness function which shows how the chromosome solves or comes close to the solution. Initial population of the genetic algorithm is created randomly. GA generates successive populations by using the following genetic operators: selection, mutation and cross over. GA obtains the optimal solution after a series of iterative computations with some termination condition.

The basic algorithm of GA is given below

Algorithm
Input:
Binary String, Number of generations, Population Size, Crossover Probability, Mutation Probability
Output: Selected subset of features
Begin
  Initialize population;
  Evaluate population members;
  While (the stopping criterion is not met)
  Begin
    Select parents from current population;
    Apply genetic operators to selected parents;
    Evaluate offspring;
    Set offspring equal to current population;
  End
    Next generation until stopping criterion
End

Initial population is evaluated using fitness function. Based on the fitness values, the parents are selected from current population. Crossover and mutation operator affects the fitness value of a chromosome. Crossover allows the exchanging of genes between two chromosomes using the one point crossover, two point crossover, or uniform crossover. In the process of mutation the genes of chromosome may be altered, i.e. in binary code genes changing genes code from 0 to 1 or vice versa.

Offspring replaces the old population using its goodness value and forms a new population in the next generation. This process operates iteratively until termination conditions satisfy.

GA feature selection uses a wrapper approach. For a data record, each value of the feature is converted into a binary gene value, 0 or 1. We produce initial population randomly where each individual contains approximately the same number of 1's and 0's on the average. The population size is 100 and the number of generation is 100. Fitness value of each individual is calculated and its goodness is evaluated. Based on the goodness value

the individuals are selected for mutation and cross over operations.

3 Problem Solution

Fig.1 shows the conceptual architecture of proposed IDS. The architecture consists of the following components: Data Collection module, Preprocessing module, Feature Selection module, Training and Classification module.

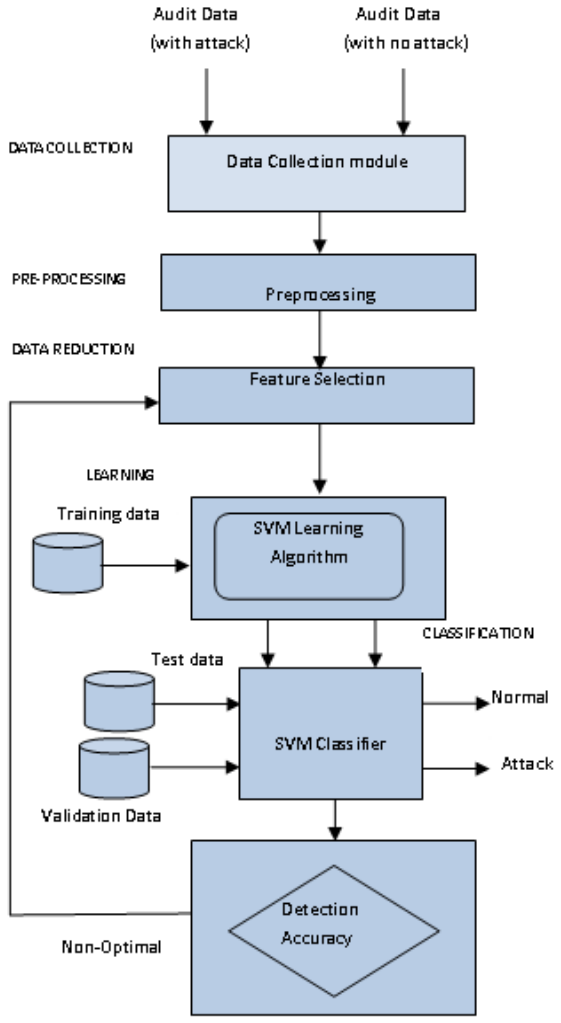

Fig.1. Architecture of the Proposed System

3.1 Data Collection Module

The data collection module collects the audit data from MAC and network layers to profile the normal and malicious behavior of mobile node. The collection module in the IDS architecture monitors the events and packet delivery time, traffic, and topology statistics and records the feature values. In anomaly detection, we want to select the trace data that bears evidence of normality or anomaly. Normal profile is created using the data collected during the normal scenario. Attack profile is created by simulating different attacks. The feature set includes routing activities and data forwarding behavior at network layer. The proposed system uses the most popular reactive Ad hoc On Demand Distance Vector Routing (AODV) [21] routing protocol for collecting the routing behavior. List of important cross layer features are shown in table 1.

Table 1 List of Cross Layer Features

Layers	Name of the Features	
MAC Layer	RTS, CTS, MAC Data Packets, ACK	
Network Layer	Routing Control Packets	RREQ, RREP, RERR and Hello
	Routing Table Packets	Number of Neighbors, Added routes, Deleted Routes etc.
	Data Control Packets	Data

3.2 Preprocessing Module

Collected data are normally incomplete and noisy. Also the values may be missing or it may contain outliers. During preprocessing these issues are handled and also collected data is transformed into appropriate format for the detection process.

3.3 Feature Selection Module

Proposed intrusion detection uses two important models for feature selection. The filter model chooses feature by using training dataset, while the wrapper model does this by the feedback of the classifier evaluation.

3.3.1 Rough Set feature Selection

Rough Set Theory uses the concept of filter approach. Feature reduction by Rough Sets Theory is carried out using ROSETTA toolkit [22]. The proposed system uses discernibility matrices for finding reducts. A discernibility matrix of a decision

table $D = (U, C \cup D)$ is symmetric. The entries are defined as

$$c_{ij} = \{a \in C \mid a(x_i) \neq a(x_j)\} i, j = 1,.....,|U|$$

Each *cij* contains those attributes that differ between objects *i* and *j*. The reduction algorithm based on discenibility matrix is outlined below.

Input: A decision table $T = (U, C \cup D)$
Output: A reduction of T, denoted as *Redu*.

Step 1: Calculate the discernibility matrix of decision table.
Step 2: For all elements of non-zero value, non-empty sets C_{ij} in the discernibility matrix, establish the relative disjunctive logic expression $L_{ij,}$

$$L_{ij} = \bigvee_{a_i \in C_{ij}} L_{ij}$$

Step 3: Operate conjunctively all the disjunctive logical expression,get a conjunctive paradigm.
Step 4: Convert conjunctive paradigm into disjunctive paradigm.
Step 5: Output attribute reduction results.

In disjunctive Paradigm each item on the conjunction corresponds to a result of attribute reduction.

3.3.2 Genetic Feature Selection

Genetic algorithm uses wrapper feature selection algorithm for the selecting the important features. Feature selection process consists of two main components: GA and SVM classifier. GA selects the subset of features and then SVM classifier evaluates the subsets during a classification process. The result of the classification is applied for finding the fitness value of GA.

Let n be the total number of features of collected data. Random individuals are created for the collected data. Each individual represent the feature sub set. SVM algorithm determines the fitness value for each individual. The fitness value depends on objective function or cost function. Fitness value for each individual is calculated and best individual is selected for next generation. GA use some type of fitness measure to evaluate the performance of each individual in a population.

Fitness(x) = accuracy(x)

Where x is the feature subset. Accuracy(x) is the detection accuracy of SVM on feature subset X. If two feature subsets have the same detection accuracy then feature subset with lowest feature is selected. The GA algorithm applied repeatedly until a defined generation is reached. Finally, it produced the optimal set of features. The reduced feature set is used for training and testing process. The algorithm for feature selection is given below.

Input: Data Set
Input for GA: number of chromosomes, total iteration, crossover and mutation rate
Output: selected subset features
Begin
Step 1: Read the total number of features from the collected data.
Step 2: Produce the random initial populations by turning on/off the individual bits.
Step 3: Evaluate each individual.
Step 3.1: Read the feature values.
Step 3.2: chromosome representation in New-GASVM.
Step 3.2.1: Save the values of each feature and store it in array.
Step 3.2.2: Sort the values.
Step 3.2.3: Store the selected features based on the values.
Step 3.3: Evaluate each individual (chromosome) using SVM classifier.
Step 4: GA operates on the population to evolve the best solution.
Step 4.1: Apply the selection strategy and GA operators.
Step 4.2: Repeat Step 3
Step 5: Return the best subset of features.
End

3.4 Training Module

Machine learning is used for the training purpose of proposed IDS. The learning model is essentially a Support Vector Machines (SVM). This model is trained by SVM algorithm using the reduced training set. Given a set of training examples, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other.

An SVM maps linear algorithms into non-linear space. It uses a feature called, kernel function, for this mapping. Kernel function is used to divide the feature space by constructing a hyperplane. The kernel functions can be used at the time of training of the classifiers which selects support vectors along the surface of this function. In this module SVM is trained by using the training data set.

3.5 Classification

The trained SVM model can be used for the detection of malicious behavior. For many problems it is not easy to find hyper planes to classify the data. The SVM has several kernel functions that users can apply to solving different problems. Selecting the appropriate kernel function can solve the problem of linear inseparability. Another important capability of the SVM is that it can deal with linear inseparable problems. Internal product operations affect the classification function. A suitable inner product function $K(x_i,x)$ can solve certain linear inseparable problems without increasing the complexity of the calculation. There are four kernel function namely, linear function, polynomial function, radial basis function (RBF), and sigmoid function.

The decision function for non linear SVM is

$$f(x) = sgn(\sum_{0}^{N_S} \alpha_i y_i k(x_i,x) + b$$

Ns is the total number of support vectors. The sign of the decision function determine the category of unknown behavior x. $K(x_i,x)$ is the kernel function. The proposed IDS use RBF as the kernel function.

4 Experimental Results

To validate the efficiency of the proposed IDS model attacks are simulated with varying network conditions under Linux environment using Network Simulator (ns-2) [23]. Table 2 lists the different values of experiment parameters. The three network conditions mobility, traffic density and number of malicious nodes are varied for analysis. There are 13 different scenarios: five in varying mobility conditions, four in varying traffic density and four in varying number of malicious nodes. Traffic density represents the number of nodes involved in the transmission. Mobility is varied by varying the pause time of the mobile nodes.

In this work one of the most popular reactive routing protocol of a MANET, Ad hoc On Demand Distance Vector (AODV) is used. AODV is designed such that all the nodes must participate in the routing process. This protocol assumes that the network is trusted and the nodes are cooperative. But AODV is vulnerable to wide variety of attacks like Route Disruption, Route Invasion, Node Isolation and Resource Consumption [24]. The trace files are generated by simulating the attacks with different mobility of a node. The features are collected by each node periodically by analyzing the data from the trace log using awk scripts. All these features are only local to the nodes.

The node mobility is varied and how mobility affects the detection rate is studied. Similarly, by varying the traffic density and number of malicious nodes is experimented and the effect of these conditions over detection efficiency is studied.

Table 2 Ns-2 Attack Simulation setup with varying network conditions

S.No	Parameter	Value
1	Routing protocol	AODV
2	Simulation duration	1000 sec
3	Topology	1000m x 500 m
4	Number of mobile nodes	50
5	Transmission range	250 m
6	Mobility model	Random waypoint model
7	Traffic type	CBR/UDP
8	Data payload	512 bytes
9	Number of connections	5,10,15 and 20
10	Maximum speed	20 m/s
11	Number of malicious nodes	5,10,15 and 20
12	Pause time	0,20,40,60 and 80
13	Attack duration	2 – 50 sec

LIBSVM tool [25] is used for the SVM operations. For each scenario, twenty five individual runs with different network conditions are performed.

To measure the performance there are three metrics used in the evaluation namely, detection accuracy, false positive rate and false negative rate. Detection accuracy (DA) is defined as the ratio of the number of events being predicted correctly to the total number of events. False positive rate (FPR) is defined as the ratio of the number of attack-free events falsely being identified as anomalies to the total number of normal events. False negative rate (FNR) is defined as the ratio of the number of anomalies falsely predicted as attack-free events to the total number of anomalies.

The results of IDS are compared between SVM aided by Rough Set and SVM aided by genetic algorithm for the Route Disruption Attack.

4.1 Effect of Mobility

These three parameters are affected by mobility. To see this consequence performance metrics are measured with five mobility levels, i.e. pause time is set to 0, 20, 40, 60 and 80. It shows false positive rate, false negative rate increases and detection accuracy increases as mobility decreases (or pause time increases).Fig. 2 depicts the relationship between detection accuracy and pause time. If the pause time is 0, the nodes are moving in the network all the time. Due to the mobility, the detection accuracy is reduced. If the pause time is increased, then the nodes are closer to static position. Therefore the detection accuracy is improved if the pause time is increased.

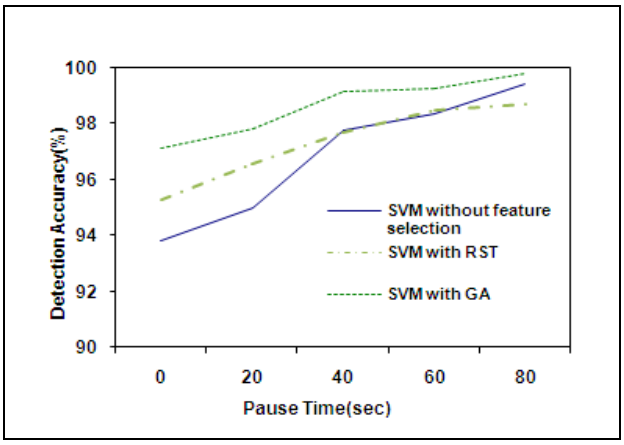

Fig.2. Pause Time vs. Detection Accuracy

Fig.3 and Fig.4 depicts the relationship between incorrect predictions (false positive rate and false negative rate) and pause time. From that we infer that, if the nodes are not moving the activities of the network can be easily predicted.

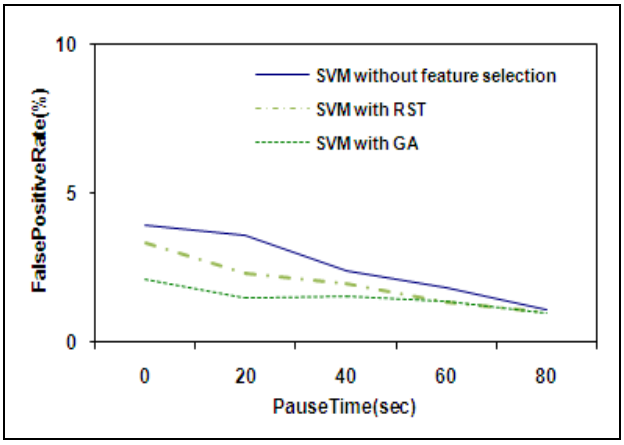

Fig.3. Pause Time vs. False Positive Rate

4.2 Effect of Traffic Density

Fig.5 illustrates the effect of network traffic against detection accuracy. Metrics are measured with different traffic levels such as 5, 10, 15, and 20 data sources. Here the experiments are done with the pause time of 80 seconds and number of malicious nodes with 5.

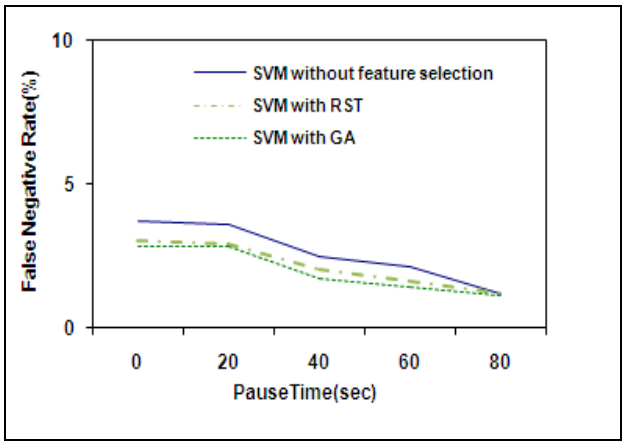

Fig.4. Pause Time vs. False Negative Rate

Observation shows that the number of connection increases then there is slight degradation in the performance. If the traffic increases, all malicious nodes are trying to send bogus routing control packets to every source. So the system has to take care of different anomalous activities. Due to this reason the detection accuracy is reduced by one percentage at each traffic level.

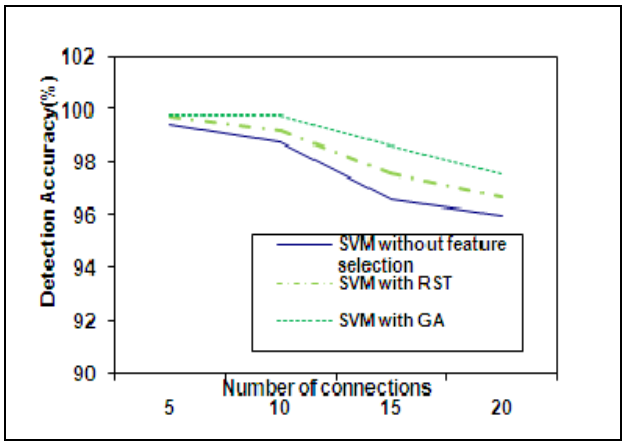

Fig.5. Number of connections vs. Detection Accuracy

4.3 Effect of Number of Malicious Nodes

Fig.6 illustrates the effect of different number of malicious nodes with detection accuracy. Metrics are measured with different attack levels such as 5, 10, 15, and 20 malicious nodes. Experiments were done with the pause time of 80 seconds and number of connections with 5. Here also there is a slight degradation in detection accuracy. If the number of malicious nodes is less, then they cannot send bogus messages to normal nodes all the times because of the communication range. If it increases the attacker can easily achieve its objective. Because of these characteristics the detection accuracy is decreased by the rate of 1%.

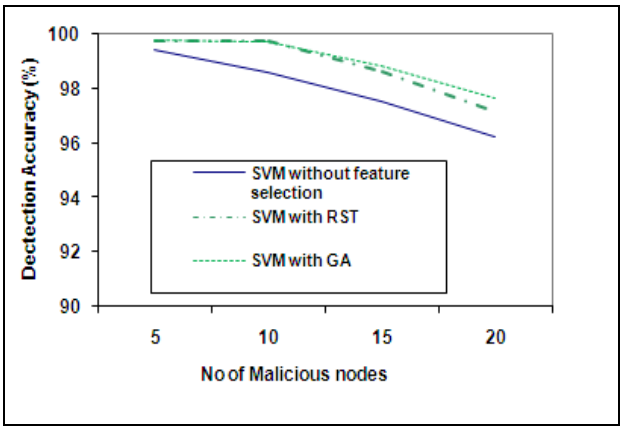

Fig.6. Number of Malicious nodes vs. Detection Accuracy

Fig. 7 summarizes the experimental results for the simulated attacks. Observation shows that the performance with feature selection is better compared with all features. Another observation is that genetic feature selection outperforms rough set feature selection.

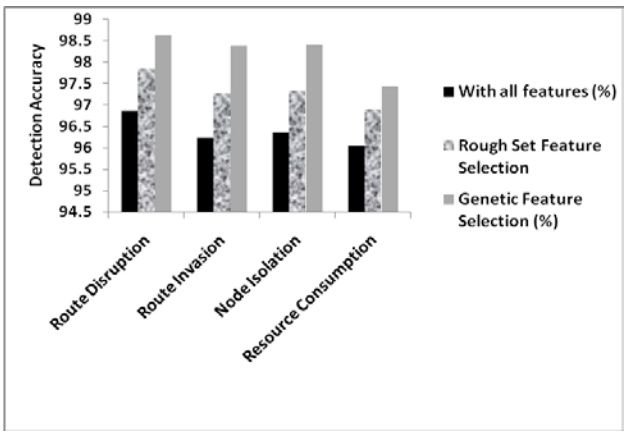

Fig.7. Experimental results for the simulated attacks

5 Conclusion

In this work, we have presented a novel approach to select the best features system for detecting misbehaviours. It uses cross layer data to characterize the behaviour of mobile nodes. Our approach uses two feature selection techniques namely, rough set theory and genetic algorithm. This approach reduces the computation overhead and increases the detection accuracy. The efficiency of the IDS is analyzed with varying network conditions by simulating routing attacks. Experimental studies shown that IDS with selected features increases the detection accuracy and minimizes the false alarm rate. The system has achieved an overall detection accuracy of detection with all features is 96.37, rough set feature selection is 97.34 and genetic feature selection is 98.22. This observation shows that the performance of IDS with feature selection is improved. Also the performance of genetic feature selection mechanism is better than that of rough set theory. This is due to the wrapper approach. Moreover GA uses support vector machine as an evaluation component, which is used to validate the selection of features at each generation.

References:

[1] Yi-an Huang and Wenke Lee, A Cooperative Intrusion Detection System for Ad Hoc Networks, *Proceedings of the 1st ACM Workshop on Security of Ad hoc and Sensor Networks, Fairfax, Virginia*, 2003, pp. 135 – 147.

[2] Mishra, A., Nadkarni, K. and Patcha, A, "Intrusion detection in wireless ad hoc networks" *IEEE Transactions on Wireless Communications*, Vol. 11, No.1, 2004, pp. 48–60.

[3] Elhadi Shakshuki, Nan Kang, Tarek R. Sheltami, "EAACK - A Secure Intrusion-Detection System for MANETs", *IEEE Transactions on Industrial Electronics*, Vol. 60, No.3,2013,pp. 1089-1098.

[4] C. F. John Felix, A. Das, B.C. Seet, and B.S. Lee, "Cross Layer versus Single Layer Approaches for Intrusion Detection in MANET," *IEEE International Conference on Networks, Adelaide*, 2007,pp. 194-199.

[5] Joseph, J.F.C., Bu-Sung Lee, Das,A. and Boon-Chong Seet, "Cross-Layer Detection of Sinking Behavior in Wireless Ad Hoc Networks Using SVM and FDA" IEEE Transactions on Dependable and Secure Computing, vol.8, no.2, 2011, pp.233-245.

[6] Khalil El-Khatib "Impact of Feature Reduction on the Efficiency of Wireless Intrusion Detection Systems" *IEEE Transactions on Parallel and Distributed Systems*, Vol. 21, No. 8, 2010,pp. 1143- 1149.

[7] X. Wang, T. L. Lin, and J. Wong, Feature Selection in Intrusion Detection System over Mobile Ad-hoc Network, Technical Report, Computer Science, Iowa State University, USA, 2005.

[8] A.H. Sung and S. Mukkamala, "The Feature Selection and Intrusion Detection Problems," *Proc. Ninth Asian Computing Science Conf.*, 2004.

[9] Neil S. Mac Parthalain, "Rough Set Extensions for Feature Selection" PhD thesis, Department of Computer Science. Aberystwyth University. Aberystwyth, United Kingdom, 2009.

[10] Rung-Ching Chen, Kai-Fan Cheng and Chia-Fen Hsieh 'Using Rough Set And Support Vector Machine for Network Intrusion Detection' *International Journal of Network Security & Its Applications (IJNSA)*,Vol 1, No 1, 2009, pp.1-13.

[11] A. Chitra and Anupriya rajkumar," Genetic Algorithm Based Feature Selection for Paraphrase Recognition" *International Journal on Artificial Intelligence Tools*, Vol.2, No.2, 2013, pp.1-17.

[12] Anand Kannan, Gerald Q. Maguire Jr., Ayush Sharma, Peter Schoo, "Genetic Algorithm Based Feature Selection Algorithm for Effective Intrusion Detection in Cloud Networks" IEEE 12th International Conference on Data Mining Workshops, 2012, pp.416-423.

[13] Sevil Sen, Zeynep Dogmus, "Feature Selection for Detection of Ad Hoc Flooding Attacks", *Advances in Intelligent and Soft Computing, Springer* 2012, Volume 176, pp 507-513.

[14] Joseph, J.F.C., Bu-Sung Lee, Das,A. and Boon-Chong Seet. "Cross-Layer Detection of Sinking Behavior in Wireless Ad Hoc Networks Using SVM and FDA" IEEE Transactions on Dependable and Secure Computing, vol.8, no.2, 2011,pp.233-245.

[15] Chih-Fong Tsai, Yu-Feng Hsu , Chia-Ying Lin and Wei-Yang Lin. "Intrusion detection by machine learning: A review" Expert Systems with Applications, 36(10), 2009,pp.11994-12000.

[16] Sergio Pastrana, Aikaterini Mitrokotsa, Agustin Orfila ,Pedro Peris-Lopez. "Evaluation of classification algorithms for intrusion detection in MANETs" Knowledge-Based Systems Vol.36, 2012, pp.217-225.

[17] Huilin Yin, Pingping Xu, and Tingting Zhu, "An Efficient Feature Redundancy Removal Approach towards Intrusion Detection in Ad Hoc Network", Proceedings of the 2009 Second International Symposium on Information Science and Engineering, pp.191-195.

[18] You Chen, Yang Li, Xue-Qi Cheng, and Li Guo, "Survey and taxonomy of feature selection algorithms in intrusion detection system" Proceedings of the Second SKLOIS conference on Information Security and Cryptology, Springer-Verlag, 2006,pp.153-167.

[19] Pawlak,.Z (1998) 'Some Issues on Rough Sets' *Transactions on Rough Sets I*, pp.1-58.

[20] Goldberg, D. E. *Genetic algorithms in search optimization and machine learning*. Addison-Wesley,1989.

[21] Perkins, C., Belding-Royer, E. and Das, S.,2003. "Ad hoc On-Demand Distance Vector (AODV) Routing", *IETF RFC 3561*.

[22] Ohrn, A., Komorowski, J ," A Rough Set Toolkit for Analysis of Data", *In Proceedings of the third Joint conference on Information Sciences*, Vol.3, USA, 1997, pp.403- 407.

[23] Ns-2: The Network Simulator, 2010. http://isi.edu/nsnam/ns/.

[24] Ning,.P and Sun,.K, "How to misuse AODV: A case study of insider attacks against mobile ad-hoc routing protocols" *In Proceedings of 4th Annual. IEEE Information Assurance Workshop*, 2003, pp. 60–67.

[25] LIBSVM -- A Library for Support Vector Machines:www.csie.ntu.edu.tw/~cjlin/libsvm/