

Contents list available at JMCS

Journal of Mathematics and Computer ScienceJournal Homepage: www.tjmcs.com

A Rough Set Based Approach to Classify Node Behavior in Mobile Ad Hoc Networks

¹Mohit Jain, ²M.P.S Bhatia

Department of Computer Engineering
Netaji Subhas Institute of Technology (N.S.I.T)
University of Delhi, New Delhi, India

¹Jain.mohit@nsionline.in, ²mps.bhatia@gmail.com

Article history:**Received: April 2014****Accepted: May 2014****Available online : June 2014**

Abstract

Mobile Ad Hoc Network are used in places where providing a network infrastructure is difficult. In Ad Hoc Network the mobile nodes are not controlled by any other controlling entity, they have unrestricted mobility and form the dynamic topology. This dynamically changing network topology of MANETs makes it vulnerable to many security related issues. There are some situations when one or more nodes in the network become selfish or malicious and tend to annihilate the capacity of the network. This research investigate the classification of good and bad nodes in the network by using the concept of rough set theory , that can be employed to generate simple rules and to remove irrelevant attributes for discerning the good nodes from bad nodes. Our experiment results reveals that the rough set based approach increases the network capacity and throughput of the network up to 98.9%

Keywords: Indiscernibility, Equivalence Class, Reducts, Malicious Nodes, Rough Sets, Crisp Sets, Decision Rules.

1. Introduction

Mobile Ad-hoc Networks (MANET) are networks composed of a set of mobile hosts with wireless transceivers. They do not rely on any infrastructure or centralized management. These networks are useful when no wired link is available such as in disaster recovery or more generally when a fast deployment is necessary. Each node acting as both a host and a router moves arbitrarily and communicates with others via multiple wireless links [1]. Intermediate nodes act as routers by forwarding data. Due to node mobility, node failures, and the dynamic characteristics of the radio channel, topology of ad-hoc network is highly dynamic. Routing protocols should adapt to such dynamism, and continue to maintain connection between the source and destination nodes in the presence of path breaks caused by mobility and/or node failures.

Each node participates in ad hoc routing protocol that allows it to discover multi hop paths through

the network to any other node. In general nodes which are not cooperative and betrayed the whole network can be of two types. Selfish nodes and malicious nodes. Malicious nodes are either faulty and do not ensue the protocol or an intentionally malicious and try to attack the system. The problems created by these nodes needs to be addresses at different multiple layers , for example using spread spectrum encoding to avoid the interference over the communication channel, using a reputation system to identify the malicious system, and subsequently avoiding and penalizing such nodes. A selfish node on the other hand is an economically rational node whose objective is to maximize his own welfare, which is defined as the benefit of its actions minus the cost of its actions. Since forwarding a message will incur a cost, a selfish node will need an incentive for doing it. [2].Due to this malicious nodes or selfish nodes the performance of the network decreases rapidly and makes it difficult to manage. Today one of the Key challenges in MANET is security issue. Many researchers have proposed the methods and techniques to resolve this issue. To our knowledge this is the first work in literature which proposed a new technique known as rough set theory to resolve this issue to much deeper extent. In addition to this, we have also analyze our data set with rough set method by using train and test scenario. We have split the data set into two disjoint parts. One becomes the training sample and other is used as a testing set for classifier constructed from the training data. We obtain classification results for test sub table and structure of classifier we have learned from training data by building the confusion matrix in fig7.

The rest of the paper is organized as follows. In section 2 related work and motivation has been discussed. In section 3 we represent the background review that is important for understanding the research paper to follow. In section 4 we discussed our proposed approach that is based on rough sets. In section 5 the analysis of our proposed approach has been given using Rough Set Exploration system (RSES). In section 6 the experimental observation has been demonstrated using NS-2. Section 7 demonstrates the performance analysis of simulation scenarios using X-graphs and finally in section 8 our research work has been concluded.

2. Related work and Motivation

There are many papers in literature discussing about intrusion detection system in ad hoc networks. Generally all the papers discuss about generic methods to detect the intrusion and anomaly detection in network , which are specific in terms of detection of particular attack or anomaly behavior of nodes, no one has veer their attention toward evidence based method which can help in the classification of nodes taking part in ad hoc network. If we able to classify the nodes behavior by deriving effective decision rules from the forensic network attributes then it will automatically help in detecting the intrusion detection as well. This motivated us to develop the analytical solution for the classification of nodes in MANETs. In [4] author proposed the improved neighbor detection algorithm for AODV protocol by using the concept of signal to noise ratio. [5] Has suggested the method to decide the number of necessary nodes by considering size and transmission range of the network. In[6], author proposed watchdog and path rater technique that improved the throughput in mobile ad hoc networks in the presence of malicious nodes that consent to forward packets but unsuccessful in doing so. Hasswa et.al [7] discusses the weakness of path rater technique. Manikopolous and Ling [8] presented the architecture for mobile ad hoc network security where an intrusion detection system runs on every node to collect the local data from its host node and neighboring node within its communication range, process raw data and periodically broadcast to its neighbors to detect the malicious nodes. In [9] Lee examined the vulnerabilities of wireless ad hoc networks, the need for intrusion detection to supplement a secure mechanism.

Rough Set methodology is described in this work for the classification of good nodes with minimum number of attributes along with decision rules generated from the data set. It's a powerful technique to generate simple rules and to remove irrelevant attributes for discerning the good nodes from bad nodes

3. Background Work

In this section we first briefly introduced about the rough sets theory with the little introduction of Information systems. After that, we will introduce a detailed discussion on information system.

A. *Rough Set Theory*

Rough set theory proposed by Pawlak (1982) is a mathematical tool that's deals with vagueness and uncertainty. Its concepts and operations are defined based on the indiscernibility relation. In this theory, a data set is represented as a table, where each row represents an event or object or an example or an entity or an element. Each column represents an attribute that can be measured for an element. [3][13]. This data table is known as Information systems. The set of all elements is known as universe It has been successfully applied in selecting attributes to improve the efficiency in deriving decision rules. In Information systems, elements that have the same value for each attribute are indiscernible and are called elementary sets. Subsets of the universe with the same value of the decision attribute are called concepts. A positive element is an element of the universe that belongs to concept. For each concept , the greatest union of elementary sets contained in the concept is called the lower approximation of the concept and the least union of elementary sets contain the concept is called the upper approximation of the concept , that are not the members of the lower approximation is called the boundary region . It provides the useful information about the role of particular attributes and their subsets and prepares the ground for representation of knowledge hidden in data by means of IF-THEN decision rules. A set is said to be rough if the boundary region is non-empty and a set is said to be crisp if the boundary region is empty.

B. *Information Systems*

An information system can be viewed as a table where each row presents an object and each column present attribute .That can be measured for each object. Basically, an information system is a pair $S = (U, A)$ where U in non empty finite set of object known as universe and A is non empty finite set of attributes such that $a: U \rightarrow V_a$ for every $a \in A$ and the set V_a is called the value set of a [3]. Information systems can be extended by the inclusion of decision attributes and information systems of this kind is Known as decision systems. A decision system is an information system of the form $S = (U, A \cup \{d\})$, where $d \notin A$.Information systems can be extended by the inclusion of decision attributes and information systems of this kind is known as decision systems. A decision system is an information system of the form $S = (U, A \cup \{d\})$, where $d \notin A$ is the decision attribute and the elements of A are called condition attributes. Normally decision attribute takes one of two possible Values but it can also take multi values. A decision system expresses almost all the knowledge about the model. Sometimes in the data table the same or indiscernible objects may be represented several times or some of the attributes may be superfluous. This can be expressed as:

$$IND(B) = \{(x, x') \in U^2 \mid \forall a \in B \ a(x) = a(x')\} \quad - (1)$$

Where $IND(B)$ is an Equivalence relation and is called B-indiscernibility relation. Rough set analysis can be done lower and upper approximations. This can be defined as follows ,

$$\text{Lower Approximation: } \underline{B}_*(x) = \{x \in U : B(x) \subseteq x\} \quad - (2)$$

$$\text{Upper Approximation: } \overline{B}^*(x) = \{x \in U : B(x) \cap X \neq \emptyset\} \quad - (3)$$

Where $B \subseteq A$ and $X \subseteq U$. We can approximate X by using only the Information contained in B by constructing the lower approximation and Upper approximation defined in (1) and (2). Due to granularity of knowledge, rough sets cannot be characterized by using Available knowledge. Therefore with every rough set we associate two crisp called its lower and upper approximation. The lower approximation of sets consists of all elements that surely belong to the set. The difference of the upper and lower approximation is a boundary region and any rough set has non empty set boundary region [3]. Rough sets can be characterized numerically by the following coefficient:

$$\alpha_B(X) = \frac{|B_*(X)|}{|\overline{B}^*(X)|} \quad - (4)$$

Where $|X|$ denote the cardinality of $X = \phi$.

If $\alpha_B(X) = 1$, the set X is crisp with respect to B and if $\alpha_B(X) < 1$, the set X is rough with respect to B.

Sometimes there are some subsets of conditional attribute that preserve the partitioning of the universe and such subsets are known as minimal reducts. Such reducts can be found with the help of discernibility matrix function which can be defined from the formula:

$$C_{ij} = \{a \in A \mid a(x_i) \neq a(x_j)\} \text{ for } i, j = 1, \dots, n$$

$$F_s(a_1^* \dots a_m^*) = \bigwedge \{ \bigvee c_{ij}^* \mid 1 \leq j \leq i \leq n, c_{ij} \neq \emptyset \} \quad (5)$$

Where $c_{ij}^* = \{a^* \mid a \in c_{ij}\}$

Also we can measure the significance of the approximate reduct and the effect on the data set after dropping that particular attribute by the following formula:

$$\alpha_{(C,D)} = 1 - \gamma(C - \{a\}, D) / \gamma(C, D) \quad (6)$$

4. Proposed Approach

The word rough means 'inexact'. The word rough set defines the non empty boundary region. A mobile ad hoc network (MANET) is a network formed by a collection of nodes that are free to move around (mobile), leave and join the network at their wish. This kind of a network is beneficiary at places where building up an infrastructure is not feasible. Due to this lack of proper infrastructure and dynamic nature of MANETs it is prone to many security issues. The optimal method for monitoring the wireless ad hoc network and to decide whether node will misbehave or perform regularly is to observe the network parameters. To classify the node behavior in wireless networks, we will consider the following issues and will apply the rough set technique to solve it:

1. Not all parameters are required for monitoring the network
2. Which network parameter will be effective for this purpose is not well defined.
3. One can eclectic the subset of network parameters that can be effective for this purpose

5. Analysis of Rough Set Method on Data Sets Using RSES

In our proposed work we have choose random data set of resources like Network Transmission range, Signal Strength, Packet Delivery Ratio, Energy consumption, Node Transmission range and packets dropped. We have use Rough Set Exploration systems (RSES) to obtain the decision rules and we will apply these rules to the network scenario containing malicious nodes to detect it. RSES is a toolkit for analysis of table data based on methods and algorithms coming from the area of rough sets [10]. We will ensue the following steps in order to implement our proposed work to detect the malicious nodes

Step1: Load data to the RSES (Figure1 and Figure2)

Step2: Find the Reducts (Figure 3)

Step3: Derive the Decision rules (Figure 4)

Step4: Use the Classifier known as Decision trees to learn from the training data set. (Figure 5, Figure6 and Figure8)

Step 5: Build the confusion matrix. (Figure 7)

Step6: Apply the derived the decision rules to detect the malicious nodes

Nodes	TR _n	SS	EC	TR _t	PDR	PD	d
x ₁	H	M	H	L	H	H	B
x ₂	M	H	L	H	L	M	G
x ₃	L	M	M	H	L	L	G
x ₄	H	L	H	H	M	H	B
x ₅	L	H	M	H	L	L	G

Table 1- Data Set

Where SS = Signal Strength, TR_n = Node Transmission Range, TR_t = Network Transmission Range, PDR= Packet Delivery Ratio which is nothing but no. of packets received/ no .of packets send , PD = Packets Dropped , EC= Energy consumption and d = decision attribute. In addition to this H stands for High, M stands for Medium , L stands for Low , B stands for Bad and G stands for Good.

3 / 7	Netw...	Signa...	Energ...	Node...	Packe...	Packe...	Decis...
0:1	Medium	High	Low	High	Low	Medium	Good
0:2	Low	Medium	Medium	High	Low	Low	Good
0:3	High	Low	High	High	Medium	High	Bad

Fig 1-. Training set of Table 1

2 / 7	Netw...	Signa...	Energ...	Node...	Packe...	Packe...	Decis...
0:1	High	Medium	High	Low	High	High	Bad
0:2	Low	High	Medium	high	Low	Low	Good

Fig2- Test set of Table 1

(1-5)	Size	Pos.Reg.	SC	Reducts
1	1	1	1	{"Network Transmission Rang..
2	1	1	1	{"Signal Strength"}
3	1	1	1	{"Energy Consumption"}
4	1	1	1	{"Packet Delivery Ratio"}
5	1	1	1	{"Packets Dropped"}

Fig3:-Reducts of Table 1

(1-12)	Match	Decision rules
1	2	("Network Transmission Range"=High)= > (Decision={...
2	1	("Network Transmission Range"=Medium)= > (Decision...
3	1	("Signal Strength"=Medium)= > (Decision={Bad[1]})
4	1	("Signal Strength"=High)= > (Decision={Good[1]})
5	1	("Signal Strength"=Low)= > (Decision={Bad[1]})
6	2	("Energy Consumption"=High)= > (Decision={Bad[2]})
7	1	("Energy Consumption"=Low)= > (Decision={Good[1]})
8	1	("Packet Delivery Ratio"=High)= > (Decision={Bad[1]})
9	1	("Packet Delivery Ratio"=Low)= > (Decision={Good[1]})
10	1	("Packet Delivery Ratio"=Medium)= > (Decision={Bad[1]})
11	2	("Packets Dropped"=High)= > (Decision={Bad[2]})
12	1	("Packets Dropped"=Medium)= > (Decision={Good[1]})

Fig4- Decision Rules

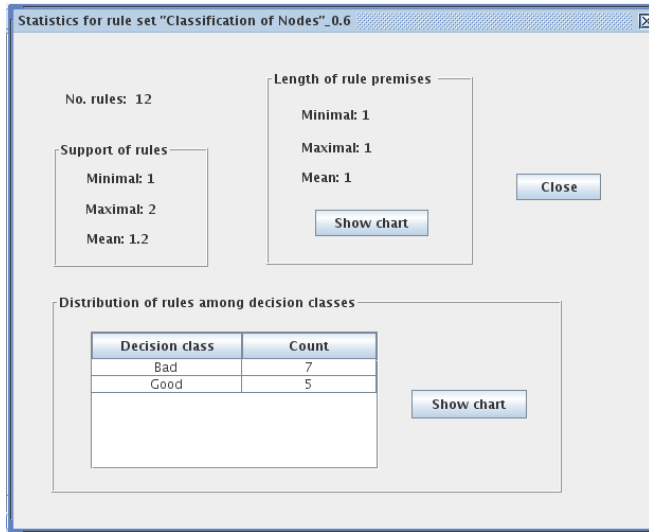


Fig5- Statistics for Rule Set

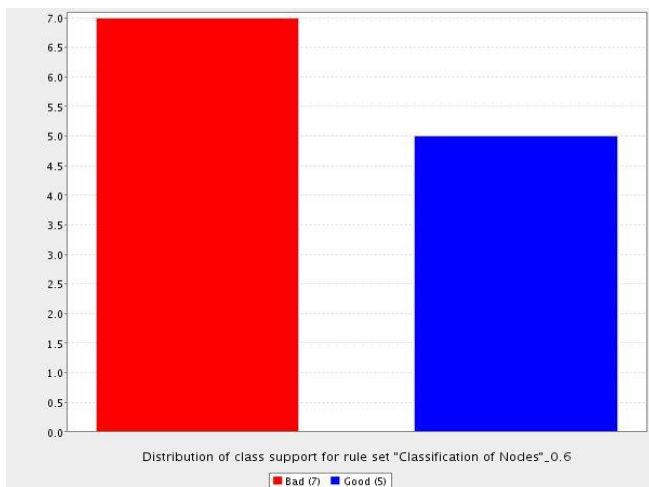


Fig6- Number of Rules Supporting Decision Classes

Results of experiments by train&test method: "Classification of Nodes"_0.4

		Predicted				
Actual		Bad	Good	No. of obj.	Accuracy	Coverage
	Bad	0	0	0	0	0
	Good	0	2	2	1	1
True positive r...		0	1			

Total number of tested objects: 2
 Total accuracy: 1
 Total coverage: 1

Fig7- Confusion Matrix

Statistics for reduct set "Classification of Nodes"_0.6

No. reducts: 5
 Size of core: 0

Stability coefficient (S...)
 Minimal: 1
 Maximal: 1
 Mean: 1

Length of reducts
 Minimal: 1
 Maximal: 1
 Mean: 1

Occurrence of attributes in reducts

Attribute	Count	Percent	Core
"Network Tran...	1	20	NO
"Signal Strength"	1	20	NO
"Energy Consu...	1	20	NO
"Packet Delive...	1	20	NO
"Packets Drop...	1	20	NO

Fig 8:-Statistics for Reduct Set

6. Implementation of Derived Decision Rules to Detect the Malicious Nodes Using NS2 Simulator

In order to classify the malicious nodes using the derived rules with the help of rough set method, we will create the network scenario of malicious nodes using NS-2 [11][12]. In the network scenario we will take different number of malicious nodes and will calculate the parameters of interest like Packet loss and packet delivery ratio. After that we will apply the decision rules to these scenarios and will calculate again the parameters like Packet Delivery ratio , packet loss and throughput in order to show that the capacity of the network increases to 98.9%. We will take the following simulation parameters:

Parameter	value
Routing Protocol	AODV
MAC Layer	IEEE 802.11

Terrain Size	500m*500m
Malicious Nodes	1,5,10
No of Nodes	20
Packet size	512 B
Initial Energy	1.5 W
R x Power consumption	0.1 W
T x Power Consumption	0.1 W
Simulation time	200 s
Traffic source	TCP

Table 2- Simulation Parameters

Scenario 1: Simulation of MANET with 1 malicious node: In the first scenario, we add the malicious behavior to Node 19. Node 19 being a malicious Node absorbs the packets in the connection among the mobile nodes. Figure 9 shows how the malicious Node absorbs the traffic.

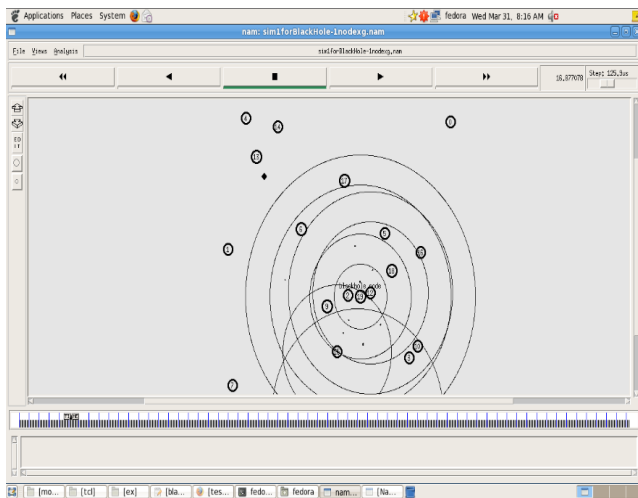


Fig 9- Scenario 1

Scenario 2: Simulation of MANET with 5 malicious nodes:

In the second scenario shown in fig 10, we add the malicious nodes behavior to Nodes 15 to 19.

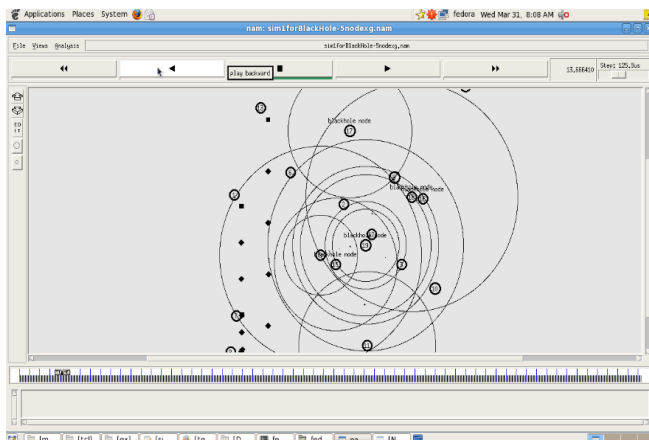


Fig 10: Second Scenario

Scenario 3: Simulation of MANET with 10 malicious nodes:

In the third scenario shown in fig 11, we could easily add the malicious behavior to Nodes 10 to 19. The malicious node attack here in this case is 50%. This also shows that malicious node Attack affects the overall network connectivity and the data loss could show the existence of the malicious nodes Attack in the network. If the number of malicious nodes is increased then the data loss would also be expected to increase.

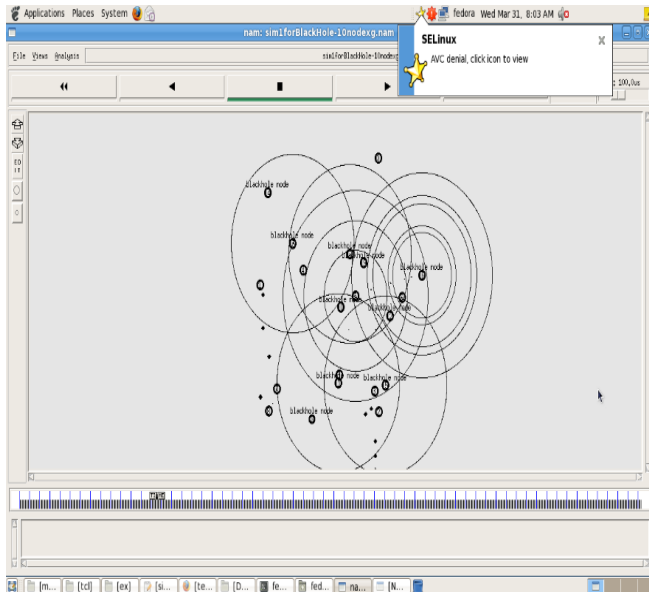


Fig 11- Third Scenario

Table 3-Simulation Results for Scenario 1:

Sending Node -> Receiving Node	Sent Packets	Received Packets	Malicious Node Drop	Loss %	Malicious node Loss %
N1->N2	1196	0	1060	100	88.62
N3->N4	3678	2190	1315	40.45	35.75
N5->N6	6019	53	4989	99.11	82.88
N7->N8	3234	3234	0	0	0

N9->N10	4471	1210	1743	72.93	38.98
N11->N12	3793	2397	105	36.80	2.76
N13>N4	3818	0	58	100	1.51
N15->N16	4663	2038	717	56.29	15.37
N17->N18	5196	189	3546	96.36	68.24

Table 4- Simulation Results for Scenario 2:

Sendin g Node -> Receivi ng Node	Sent Packe ts	Receiv ed Packet s	Malicio us Node Drop	Loss %	Malicio us Node Loss %
N1->N2	2057	0	1491	100	72.48
N3->N4	5691	2193	1020	61. 46	17.92
N5->N6	5323	4	1883	99. 92	35.37
N7->N8	4411	4358	0	1.2 0	0
N9->N10	5450	642	1901	88. 22	34.88
N11->N12	5051	131	864	97. 40	17.10
N13->N14	2990	17	0	99. 43	0
N15->N16	4872	4872	0	0	0
N17->N18	3374	3374	0	0	0

Table 5- Simulation Results for Scenario 3:

Sending Node -> Receiving Node	Sent Packets	Received Packets	Malicious Node Drop	Loss %	Malicious Node Loss %
N1->N2	4209	13	1043	99.69	24.78
N3->N4	4409	11	797	99.75	18.07
N5->N6	4528	0	0	100	0
N7->N8	5612	5600	0	0.21	0
N9->N10	5545	989	0	82.16	0
N11->N12	4434	4434	0	0	0
N13->N14	5068	5068	0	0	0
N15->N16	5352	5352	0	0	0
N17->N18	2914	2914	0	0	0

In the tables, the second column shows how many packets are sent by sending nodes and the third column shows how many of them reached the receiving nodes. By calculating the difference between the tables of normal and malicious node in AODV network we try to evaluate how many of the packets which could not reach the destination node are absorbed in the malicious Node. Packets lost in the malicious Node are shown in the fourth column of the table of the malicious node network. The rest of the columns show percentage of the packets lost and additionally in the table of malicious nodes network, we added percentage of loss packets which are absorbed in the malicious Node.

Scenario 4: In This scenario we have applied the decision rules described in fig5 to the above scenarios and calculate the network parameters packet loss and packet delivery ratio again., shown in fig12.

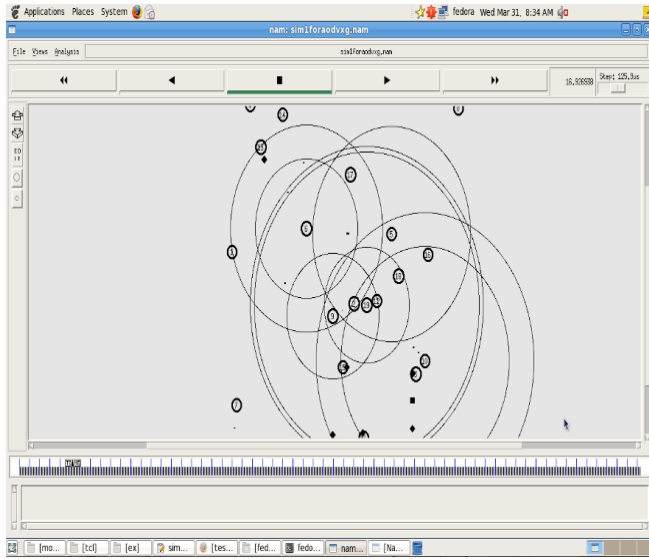


Fig12- Fourth Scenario.

Table 6- Simulation Results for Scenario 4:

Sending Node -> Receiving Node	Sent Packets	Received Packets	Malicious Node Drop	Loss %	Malicious Node Loss %
N1->N2	1139	454	0	60.14	0
N3->N4	4876	4517	0	7.36	0
N5->N6	3604	3403	0	5.57	0
N7->N8	4849	4849	0	0	0
N9->N10	3739	2978	0	20.35	0
N11->N12	2210	2210	0	0	0
N13->N14	2312	2258	0	2.33	0
N15->N16	4652	4377	0	5.91	0
N17->N18	4466	3427	0	23.26	0

Generated Packets	8532
Received Packets	8444
Forward Packets	3473
Packets Dropped	68
Average end to end delay	80.9977 ms

Packet Delivery Ratio	98.9646
-----------------------	---------

Table 7- Packet Delivery Ratio Reaches to 98.9% in Fourth Scenario

7. Performance Analysis of Scenarios Using X-Graphs

Examination of Performance for Scenario 1 and Scenario 4

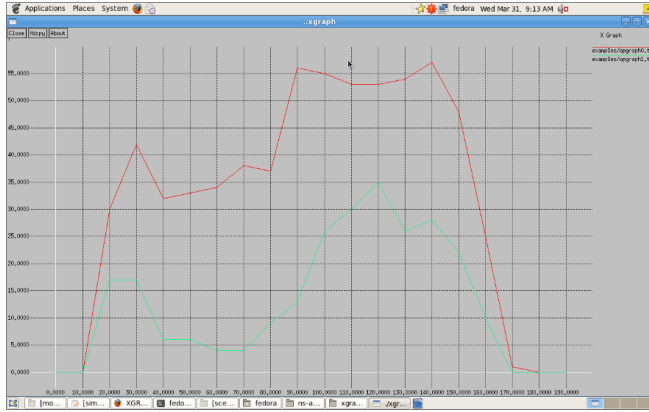


Fig 13 (Throughput v/s time for scenario 1 and scenario 4)

Examination of Performance for Scenario 2 and Scenario 4

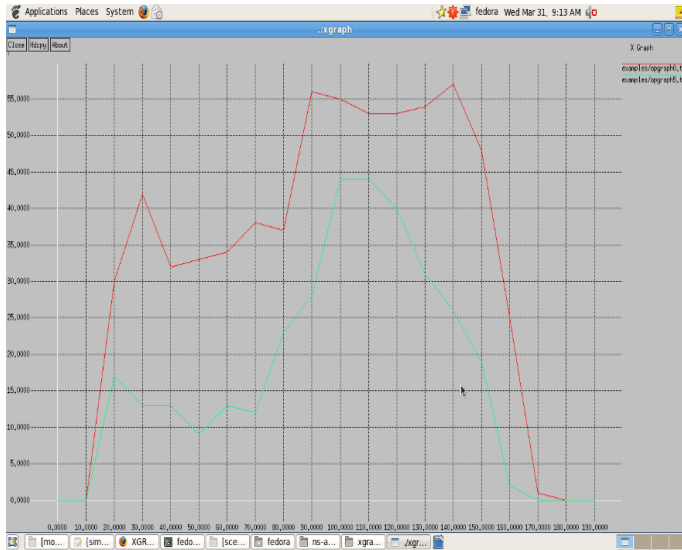


Fig 14- Throughput v/s time for scenario 2 and scenario 4

Examination of Performance for Scenario 3 and Scenario 4:

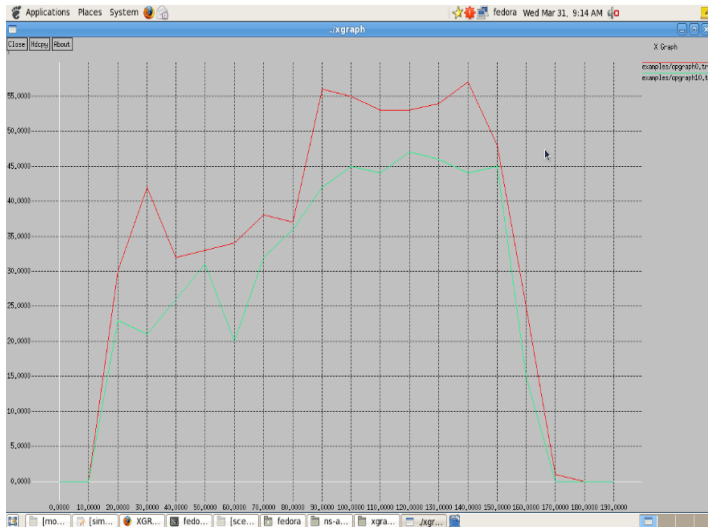


Fig 15- Throughput v/s time for scenario 3 and scenario 4

8. Conclusion

Thus the rough set approach for the classifying the node behavior has been proposed in this paper. Rough set method helps in removing the superfluous attributes and gives the minimal set of attributes known as reducts by preserving the partition of the universe of discourse and generate the decision rules. We have created the different network scenarios of malicious nodes and applied the decision rules derived from the data set to these scenarios and have show that the network capacity increases and packet delivery ratio reaches to 98.9%. In addition to this we have also analyze our data set using rough set method and evaluate the classifier by constructing the confusion matrix. We hope that our proposed approach will prove beneficiary for the ad hoc practitioners and scientist in developing the model and design for ad hoc networks.

11. References

- [1] Ehsan Hemmati and Mansour Sheikhan, “Reliable disjoint path selection in Mobile ad hoc network using noisy hop field neural network “, 2010 5th International symposium on Telecommunications.
- [2]. Uman Singh, Prof B.V.R. Reddy and Prof M.N.Hoda.” GNDA: Detecting good neighbor nodes in ad hoc routing protocol “, 2011 Second International Conference on Emerging Applications of Information Technology.
- [3]. Jan Komorowski, Lech Polkowski and Andrej Skowron, “Rough Sets: A Tu
- [4]. Srdjan Krco and Marina Dupcinov, “Improved Neighbor Detection Algorithm for AODV Routing Protocol”IEEECOMMUNICATIONS LETTERS, VOL 7, NO. 12, DECEMBER 2003.
- [5] Youngrag Kim, Shuhrat Dehkanov, Heejoo Park, Jaeil Kim, Chonggun Kim, “The Number of Necessary Nodes for Ad Hoc Network Areas ”, 2007 IEEE Asia-Pacific Services Computing Conference
- [6] S. Marti, T.J. Giuli, K. Lai, M. Baker: *Mitigating Routing Misbehavior in Mobile Ad Hoc Networks*, in 6th International Conference on Mobile computing and Networking, MOBICOM’00, P255-265, Aug 2000
- [7] A. Hasswa, M. Zulker, and H. Hassanein, *Route guard: an intrusion detection and response system for mobile ad hoc networks*, Wireless and Mobile Computing, Networking and Communication 2005, P336-343, Vol. 3, August 2005.

- [8] C. Manikopoulos, Li Ling: *Architecture of the mobile ad hoc network security (MANS) system*, in: Proceedings of the IEEE International conference on Systems, Man and Cybernetics, vol. 4, October 2003, pp. 3122-3127.
- [9] Y Zhang, W. Lee: *Intrusion Detection in Wireless ad hoc Networks*, Mobicom 2000, August 6-11, 2000, Boston, Massachusetts, USA
- [10]. RSES Homepage <http://logic.mimuw.edu.pl/>rses>
- [11] NS by Example, <http://nile.wpi.edu/NS/overview.html>, 14 May 2006.
- [12] K Fall and K. Varadhan, The NS Manual, November 18, 2005, http://www.isi.edu/nsnam/ns/doc/ns_doc.pdf. 25 July 2005.
- [13] Riki.M ,Rezaei.H , Introduction of rough set theory and Application in Data analysis , Journal of Mathematics and Computer Science 9 (2014), 25-32.

12. Copyright forms

You must include your fully-completed, signed JMCS copyright release form when you submit your paper. WE MUST HAVE THIS FORM BEFORE YOUR PAPER CAN BE PUBLISHED IN THE PROCEEDINGS. The copyright form is available from journal home page.