University of Piraeus

Department of Digital Systems

Postgraduate Program: Digital Systems Security

An experimental analysis of AODV based on sequence numbers to detect Blackhole attacks

Master Thesis

Stefanos Malliaros

Supervisor: Assistant Professor Dr. Christos Xenakis

Piraeus May 2013

Abstract

Ένας ερευνητικός τομέας ο οποίος γνωρίζει μεγάλη άνθηση τα τελευταία χρόνια είναι τα ασύρματα δίκτυα χωρίς υποδομές, ή αλλιώς Ad hoc Networks. Ιδιαίτερο ενδιαφέρον υπάρχει στην περίπτωση που οι κόμβοι των δικτύων αυτών κινούνται (MANETs – Mobile Ad Hoc Networks).

Τα δίκτυα αυτά παρουσιάζουν ένα μεγάλο αριθμό αδυναμιών καθιστώντας τα ευάλωτα σε επιθέσεις ασφαλείας. Στην παρούσα διπλωματική εργασία μελετήθηκε πειραματικά ο πιο γνωστός αλγόριθμος δρομολόγησης για αυτά τα δίκτυα Ad hoc On Demand Distance Vector (AODV). Στην συνέχεια μελετήθηκε η επίθεση Black hole κατά την οποία ένας κακόβουλος κόμβος εξαναγκάζει τους κόμβους να του προωθούν δεδομένα και αυτός τα απορρίπτει. Τέλος, με βάση τα ευρήματα από τις εξομοιώσεις που πραγματοποιήθηκαν, κατασκευάστηκε μία μηχανή ανίχνευσης κακόβουλων κόμβων οι οποίοι πραγματοποιούν την Black Hole επίθεση. Για την βάση αυτής της μηχανής χρησιμοποιήθηκε ο αλγόριθμος CUSUM.

Keywords: AODV, MANETS, ADHOC, BLACKHOLE, CUSUM, IDS

Acknowledgements

.

First of all, I would strongly like to thank my advisor, Dr. Christos Xenakis, as well as Dr. Christoforos Dadoyan for the inspiration, the guidance, the support and their trust for the writing of this thesis. I would also like to express my gratitude for necessary the technical infrastructure they provided for the completeness for this thesis. Without their continuous support I wouldn’t have the strength to continue up to the end.

Furthermore, I would like to thank all of my colleagues for the two wonderful years of this postgraduate program.

Lastly, I would like to thank my family and Chara Michalatou for supporting me throughout the difficult moments I passed during these two years.

Table of Contents

Table of Figures

An experimental analysis of AODV based on Sequence Numbers to detect Blackhole Attacks

Abstract: Mobile ad hoc networks (MANETs) are one of the recent fields that is of great interest because of their self-configuration and self-maintenance. Although they have several advantages over wired networks, they are more vulnerable to attacks in comparison to wired networks. One of the widely used routing protocol for MANETs, is the Ad hoc On Demand Distance Vector (AODV). This protocol uses sequence numbers to avoid replay attacks and routing problems. There are a lot of security attacks in ad hoc networks which can be classified into several layers. One of the most important layers is the routing layer, and one attack that belongs to it is the blackhole attack. A malicious blackhole node acts as a sink hole sucking all data packets and discarding them. In this paper, there is an experimental analysis of AODV based on the sequence numbers to detect Black Hole attacks. There is an extensive analysis on how the sequence numbers behave and based on the analysis, we show the difference of the malicious black hole behavior and the normal one. Moreover, based on the behavior of the sequence number, a detection mechanism is proposed which is based on the Cumulative Sum (CUSUM) algorithm.

1. Introduction

Mobile ad hoc networks (MANETs) are one of the recent fields that is of great interest because of their self-configuration and self-maintenance. Although MANETs have several advantages over wired networks, they are more vulnerable in comparison to wired networks. Due to their unpredictable topology, wireless shared medium, heterogeneous resources and stringent resource constraints, MANETs are vulnerable to a variety of attacks.

In MANETs, the nodes by themselves communicate with each other creating a dynamic network topology. Routing protocols play a significant role in the creation and maintenance of this topology. In MANETs, every node participates in the routing process by forwarding data to the other nodes. Many different types of routing protocols have been developed for ad hoc networks, which most of them can be classified into two main categories. The first one contains the proactive (periodic) protocols and the second the reactive (on demand) protocols. In the proactive protocols, nodes periodically exchange routing information in an attempt to form a routing table containing all the possible destinations of the network, while in reactive protocols nodes exchange routing information only when needed. As a result routing information, in the latter category, will be transmitted only when there is a need of communication. A routing protocol that belongs in the second category is the Ad hoc On-Demand Distance Vector (AODV) routing protocol. It offers quick adaptation to dynamic link conditions, low processing and memory overhead, low network utilization, and determines unicast routes to destinations within the ad hoc network.

MANETs are vulnerable to a wide variety of attacks which can be classified into several layers. One of the most important layers is the routing layer, and one attack that belongs to it, is the blackhole attack. The blackhole attack belongs to the routing disruption attacks category. A malicious blackhole node acts as a sink hole for all data packets by pretending to have a valid path for every node in the network. However, the malicious node instead of forwarding the packets, it discards them.

The AODV protocol was not designed to be secure. Therefore researchers have proposed detection techniques for several security attacks. However, the unstable topology of the network makes many of the proposed solutions to work inefficiently. Since MANETs have a wide range of applicability, the security measures need to work efficiently for networks of any size.

This paper is the first that studies the behavior of the sequence numbers of the AODV protocol. This behavior is similar independently from the size and the mobility of the mobile ad hoc network.Moreover, this paper compares the normal behavior of the sequence numbers to the abnormal behavior of a network which is under the blackhole attack. Since this behavior is the same independently from the size of the network, the sequence numbers can be used to detect efficiently a blackhole attack regardless the mobility or the size of the network. As a result, an Intrusion Detection System is proposed which is based on the nonparametric receursive version of the Cumulative Sum (CUSUM) algorithm.

The sections below are organized as follows. Section 2 discusses the necessary background and the related work. In section 3 the simulator used in this paper is presented, and the behavior of the Sequence numbers of AODV is explained. Section 4 presentsthe abnormal behavior of the sequence numbers in networks under the blackhole attack. In section 5 an IDS is proposed based on the results of the sections 3 and 4. Finally, section 6 presents the conclusions of this work and also addresses the future work that needs to be done.

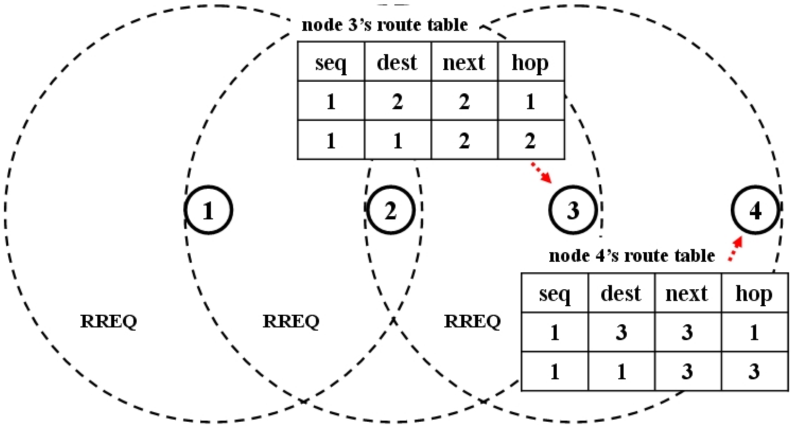
2. Background

Ad hoc On Demand Distance Vector (AODV) routing [15] algorithm is a protocol designed for mobile ad hoc networks. It is an on demand algorithm meaning that it creates routes between nodes at the start of communication. Sequence numbers are used by AODV in order to avoid routing problems, replay attacks and also to ensure the freshness of the routes. Every node increases its own sequence number when the links change. Nodes decide whether the information received is new according to the sequence numbers.

Route creation in AODV is a two step process. In the first step the source nodes searches for the destination by sending a Route Request (RREQ) message. This message is forwarded by the other nodes,using the flooding procedure. Upon reaching its' destination or a node that has an active route to the destination, a Route Reply (RREP) message is created and is sent to the source node using the reverse path. If multiple RREPs arrive at the source, the source node compares the sequence numbers and accepts the RREP with the higher one. In case the Sequence numbers are the same it accepts the RREP with the smallest number of hops till the destination.Whenever a node realizes a disconnection of the route, it creates a Route Error (RERR) message. This message is sent to the source node informing all the nodes in the reverse path about the disconnection issue. In such a case the route creation process starts from the beginning.

One of the most important parts of AODV is the use of sequence numbers. They are used in order to avoid the routing problem, replay attacks and to determine the freshness of the information. The sequence number is an increasing counter and every node needs to maintain its' own. Every node is responsible of increasing its' sequence number and inserting the newly generated sequence numbers in the control packets so that each control packet has a unique sequence number. Consequently if a node receives two packets, from the same node, with different sequence number,it knows that the packet with the higher sequence number was generated last and as a result it is considered fresher.

In AODV each node has a routing table in which all the necessary information for the routing process is stored. The routing table consists of several columns. The most important are the columns that that contain the information about the Destination node, the next hop node, the hop count till the destination and the last known sequence number for the destination node. In the below figure an ad hoc network of three nodes can be seen and their routing tables.



node 3's route table

seq	dest	next	hop
1	2	2	1
1	1	2	2

node 4's route table

seq	dest	next	hop
1	3	3	1
1	1	3	3

Figure 1: Routing tables of nodes using AODV

Blackhole is an attack that can be classified in the Routing Disruption category. A malicious blackhole node acts as a sink hole for all data packets. In order to achieve this, the malicious node replies to the received RREQs, without searching its' routing table, with a fake RREP. In order for the fake RREP to be accepted by the source it needs to have a higher sequence number than the one the source knows for the destination. Consequently, the attacker enters a large sequence number in order to be accepted for the routing process. Moreover, the highest the sequence number is, the longer the attacker participates in the network routes carrying out the blackhole attack. Eventually, the source node will trust the attacker due to the fake sequence number and it will forward all the data to the malicious node. The blackhole node discards all the data interrupting the communication between the source and the destination.

2.1 Related Work

Currently a few detection techniques have been proposed to detect the black hole attack by monitoring the AODV operation. Firstly, Detection, Prevention and Reactive Ad hoc On-Demand Distance Vector (DPRAODV) routing [1] creates a network state by measuring the number of RREQs, the number of RREPs, and the average distance of the incoming sequence numbers. Using the collected data, it calulates a threshold value and compares the next network state with it. In this solution a sudden increase in the mobility will cause a high increase in the sequence numbers which will lead to a high false positive rate.

In [2] the authors proposed a modified version of AODV in order to identify multiple blackhole nodes cooperating as a group. Each node maintains a table named Data Routing Information (DRI). In this table each node stores two bits for every node in the network. These bits help the nodes know if they have ever received or forwarded any data through the corresponding node. In this protocol the node responding to the RREQ has to send the Next Hop Node (NHN) and its' DRI entry for NHN. If the source node hasn't routed any data through the node that generated the RREP, it will ask the NHN to verify that it has a valid path to the destination. This procedure causes a considerable delay and the higher the mobility is, the chosen path from the source node might not be available when the procedure finishes. Furthermore this solution makes the routing process more complex since the nodes need to maintain extra information and exchange more packets than necessary.

The authors in [3] proposed two different approaches to solve the black hole attack in MANETs. In the first one the source nodes need to verify the authenticity of the node which sends the RREP by utilizing the redundancy (many possible paths) of the network. In this solution The RREPs contain all the hops of the path and when multiple RREPs arrive at the source, the source compares the paths and takes a decision. However, the delay caused by this procedure is critical because the selected route by the source might not be available. Also the complexity of the comparison of the routes and the data which need to be stored should be taken under consideration. In the second solution, nodes need to maintain two small tables. They have to store the last sequence number that was either sent to, or received from the other nodes respectively. They insert in the packets they send the corresponding sequence number and the nodes that receive it check if it matches with the one they have already stored in their tables. If there is a mismatch then the node that sent the packet is considered to be malicious. In high mobility scenarios packets will be dropped due to mobility issues. When the next packet is sent, with a new sequence number, a mismatch will occur resulting in a false positive ALARM signal.

In [4] H.Deng, W. Li and D.P. Agrawal proposed a solution for single black hole node detection. In this approach, the source node tries to ensure that there is a next hop node after the node that initiates the RREP. The source extracts the NHN from the RREP and sends a check packet to the NHN to verify that it has a valid path to the destination. This procedure might take enough time and according to the speed of nodes the path that the source will choose might not be available making the communication infeasible.

In [5] the authors proposed a distributed and cooperative procedure to detect malicious black hole nodes. This approach consists of four steps. In the first phase, every node constructs an estimation table by overhearing packets from the network. If a node is supposed to be malicious, the detection phases begin. These phases consist of local and cooperative detection. The final phase is the information of nodes about the malicious one. In this solution the mobility has significant effects on the detection phases. The nodes that participate in the detection must be one hop neighbors. Since the detection processes consists of several steps these nodes might not be neighbors when the detection decision should be taken. Consequently, the higher the mobility is, the higher the false positive rate will be.

Finally, Sun te al. [6],[7] has proposed a cooperative IDS architecture that focuses on routing disruption attacks using an intrusion detection engine based on statistical methods with adjustable threshold values. This use of thresholds ensures that periodical changes in routing information, caused by nodes mobility, remain under the detection threshold, while malicious behaviors that are persistent exceed the threshold indicating the occurrence of attack. However with the increase of the mobility an increase in the false positive rate and a decrease in the detection rate are also observed.

To sum up, the blackhole attack is a problem which hasn't been solved completely. The increase of the mobility causes the network topology to change rapidly and this affects the performance of the already proposed solutions. Furthermore, some of the solutions are complex while others cause delay in the communication. Some other solutions require significant changes in the existing routing protocols which makes their usage suboptimal. Based on the above, the solution the solution to the blackhole attack has to be simple and computationally easy in order to run properly in a quickly changing environment, and it should not require changes in the existing routing protocols and thus making its’ adoption easy and smooth.

In this paper, we study the behavior of the sequence numbers of AODV and based on this we propose an IDS based on the CUSUM algorithm to detect blackhole attacks. The behavior of the sequence numbers is similar independently from the mobility and the size of the ad hoc network. Consequently, the proposed solution is suitable for every mobile ad hoc network independently from its' characteristics.

3. Sequence Number Behavior

3.1 Simulator

The simulations of this paper were done using the Network Simulator v2.35 (NS2)[16] and we also used the bonnmotion 1.5a[17] in order to create the mobility patterns. The simulation area is a square and its' dimensions are 1000x1000 m^2. In the simulation scenarios the nodes communicate in pairs exchanging data packets. Each pair consists of a source and a destination node and is called a network flow.The properties of the network flows are static. The source nodes sends packets with a rate of 5 packets/sec and the size of each is 512 bytes. The UDP protocol was used for the transmission of the data and lastly the nodes have a transmission range of 250 m.

Ad hoc networks of different size have been simulated and three different variables were used in order to describe the simulated networks. These variables are **a)** Number of Nodes**b)** Speed of nodes and **c)** Number of Network Flows. The number of nodes in the simulation varies from 10 to 40 and their maximum speed also varies from 5 to 20 m/s while their minimum speed is zero. Lastly, the values of the number of flows varies from 2 to 40.

Another important characteristic for the networks simulated is the mobility pattern of the nodes. For the mobility patterns the **a)** Random Waypoint, **b)** Random Direction and **c)** Manhattan Grid model were used. One of the most frequently used mobility models in MANET simulations is the Random Waypoint [13]. In this model nodes move in random directions with the value of velocity is uniformly distributed between [U_{min},U_{max}]. In the beginning, each node is placed randomly in the simulation terrain and starts moving to a random position with a random chosen speed. Once the nodes arrives at the destination point, it stops moving (pause time) for a period of time. After the pause time ends, the same procedure is repeated until the end of the simulation.

In the below figure a moving pattern is seen of a node moving, using the Random Wayoint model. It is seen than the node starts from the point (630,580) and the rest of the dots represent the points in which the node stops and decides the next destination point.

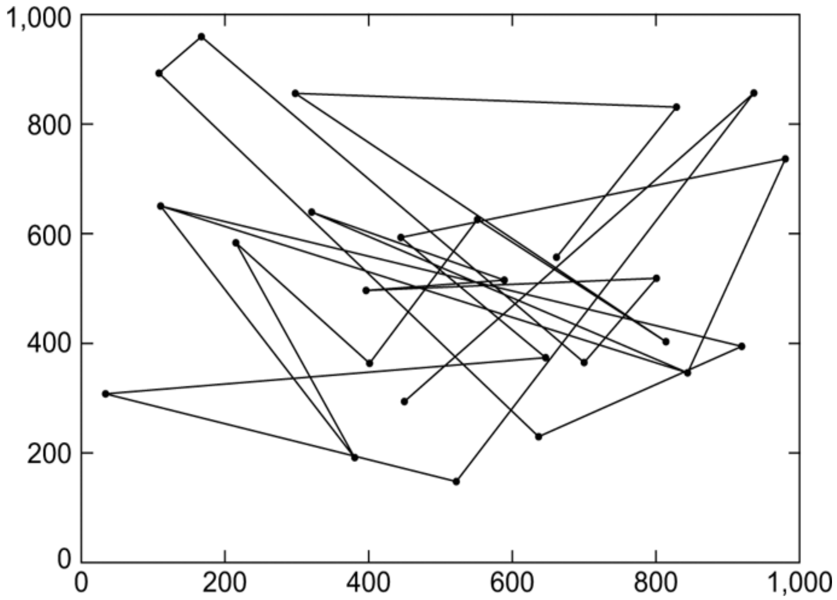


Figure 2: A moving pattern of a node which follows the Random Waypoint mobility model

Random Direction, which is a variant of the Random Waypoint Model, was first introduced in [11,12]. In this mobility model, the nodes are placed random into the terrain and choose a random direction that is uniformly distributed between [0,2π] and move towards it, until it arrives at the boundary of the simulation area. After the nodes reaches the boundaries of the simulation area, they pause for a period of time, and after that they choose another random direction between [0,π] degrees and continues the process. The value of speed is uniformly distributed between [U_{min},U_{max}].

In the next figure a node is observed moving using the Random Direction Mobility model. The node starts from the point (150,300), chooses a random direction and moves till the boundaries of the simulation area. Afterwards the node chooses a new direction and speed and start moving till the boundaries of the simulation area.

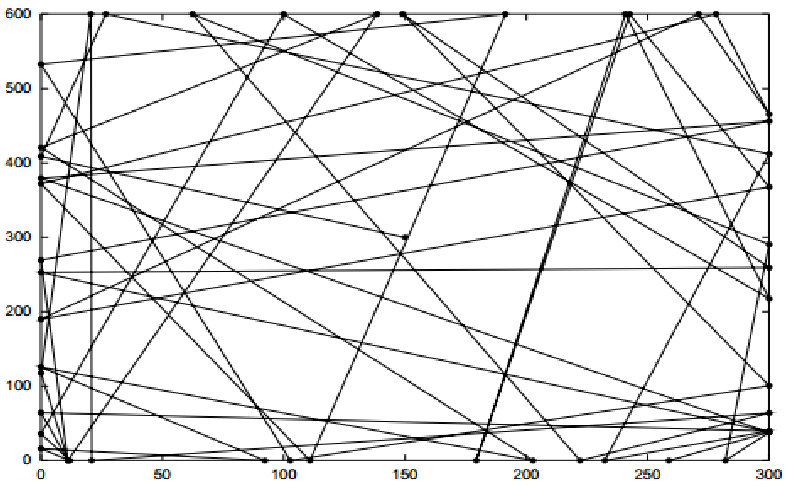


Figure 6: Traveling pattern of an MN using the Random Direction Mobility Model.

Figure 3: A moving pattern of a node which follows the Random Direction mobility model

Manhattan Grid is a more realistic mobility model than the Random Waypoint and Random Direction. It was created in order to simulate the movement of mobile nodes in an urban environment. The nodes move on horizontal and vertical lines that represent bidirectional roads of a city. The nodes can move on those roads and when they reach a crossroad, they can either continue straight ahead or turn.

In the below figure a movement pattern of a node which uses the Manhattan Grid mobility model is shown. It is clearly seen that the node moves in parallel vertical and horizontal lines. In general in this model the node has a certain possibility of moving either straight or turning when he reaches a crossroad.

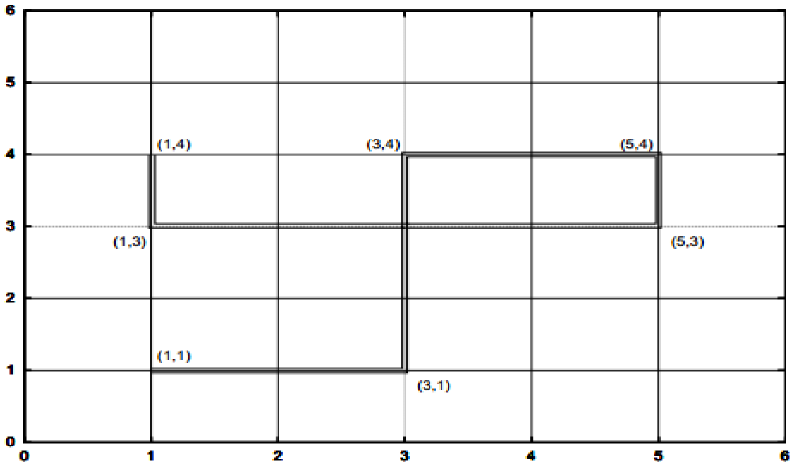


Figure 13: Traveling pattern of an MN using the City Section Mobility Model.

Figure 4: A moving pattern of a node which follows the Manhattan Grid Mobility model

To sum up a large amount of simulation scenarios have been performed. The simulation parameters described above can be seen aggregated in the below figure.

Simulation Parameters	
Number of Nodes	10-20-30-40
Min Speed / Max Speed	0 / 5-10-15-20 m/s
Simulation Terrain	1000x1000 m^2
Mobility models	RW-RD-MG
Number of Network Flows	2-40
Transmission range	250 m
Characteristics of Network Flows	
Packet rate	5 packets/sec
Packet size	512 bytes
Transmission Protocol	UDP

3.2 Simulation Results

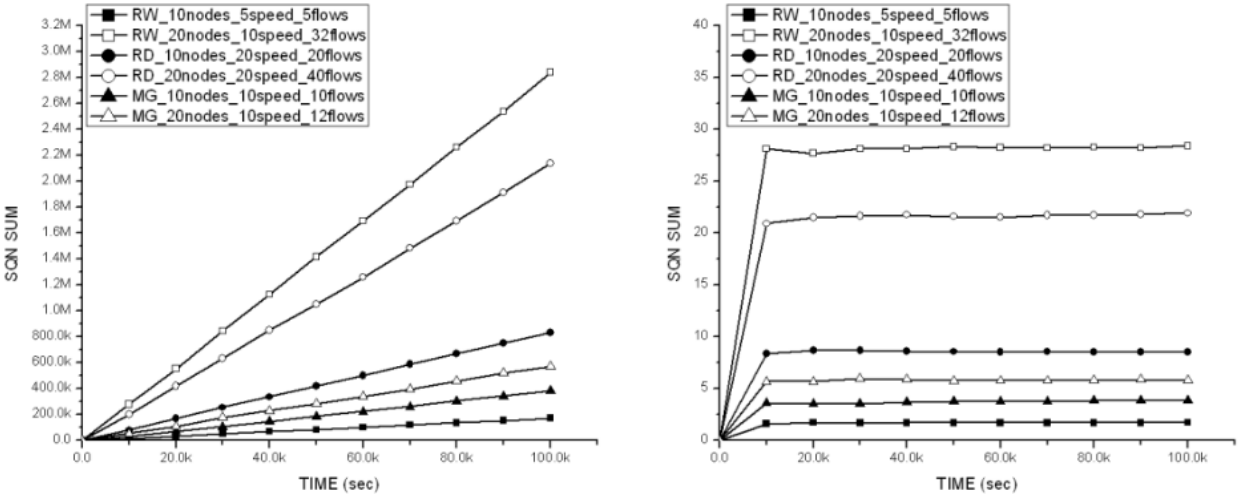
3.2.1 SQN Behavior

We have chosen to study the sum of the Sequence Numbers of the whole network. This means that in certain time intervals we sum the sequence number of every node. This choice was done because the sum of SQNs gives insights about the behavior of SQNs within the network and protocol operation, avoiding sharp variations that cause the dynamic network topology and nodes' behavior. Furthermore, it is computationally easy to calculate the summation. The dynamic topology of MANETs and the processing constraints of the nodes make more complex math operations to work inefficiently. Also the sum function can be designed as an additional module which means that no changes to the AODV protocol need to be done.

Let us see a quick example of how the sequence numbers change with an example.Suppose that we have a stationary ad hoc network with 3 nodes A, B and C. Nodes A and C have only in their range Node B. Their current sequence number are 7, 15 and 4 for nodes A, B and C respectively.Assuming that node A wants to start a communication with node C. Node A increases his sequence number to 8 because he creates a new control packet and broadcasts a Route REQuest packet (RREQ) searching node C. Inside the RREQ there are contained the Source address and Source Sequence Number (A and 8 respectively) and Destination Address and Destination Sequence Number (C and Unknown respectively). Also the hop counter in the RREQ is 1. Node B receives the RREQ and updates the information in its routing table for node A. This means that it creates an entry with the following information, Destination node A, sequence number 8, Hop Count 1 and Next Hop Node A.After that, Node B increases the hop count in the RREQ Packet to 2 and rebroadcasts the Packet with its' original Source/Destination address and sequence numbers. Node C receives the RREQ and since it is the destination node of the RREQ it has to reply with a Route REPly (RREP). Before doing so, it updates the entry for node A in its' routing table with the following information, Destination node A, sequence number 8, Hop Count 2 and Next Hop Node B. After that, node C increases its' own sequence number to 5 and replies to the RREQ with a RREP following the reverse path (C→B→A). Inside the RREP the information contained is only the Destination address of the route, which is C, Origination address of the route, which is A, andthe Destination's Sequence number, which is 5 for node C.Node B receives the RREP and updates its entry in the routing table for node C to Destination node C, sequence number 5, Hop Count 1 and Next Hop Node C.Node B will forward the RREP to node A and node A will receive the information and will update the routing information is has for node C to, Destination node C, sequence number 5, Hop Count 2 and Next Hop Node B. At this point the creation of the route has been completed and the communication may begin.

At this point, we are going to define how we compute the SQN sum. Assuming that there is a network of N nodes and SQN_i is the current SQN for node i. The SUM of the SQNs at time T of a simulation, can be calculated as $SQN_{network}=\sum SQN_i$. Assuming that, in the above example, T_1 is the time just before node A sends the first RREQ message and T_2 is the time in which node A has received the RREP, the sum for T_1 and T_2 can be calculated respectively as $SUM_{T1}=7+15+4=26$ and $SUM_{T2}=8+15+5=28$. The first significant observation is that the sumincreases linearly over time. This happens for all the chosen mobility models and independently of the size of the network.Consequently we can conclude that the rate of the sum of the sequence numbers is stabilized after a certain period of time. In our simulations we measured the rate from the beginning of the simulations. This was done using the following equation $SQN_{network\_rate}=\sum(SQN_i)/T$ where T is the simulation time.For the previous example, at Time $T_1$the rate can be calculated as $26/T_1$ and at time T_2 as $28/T_2$.

In the below figure the $SQN_{network}$ and the $SQN_{netwowrk_rate}$are shown for different networks. It is seen that the sum (Fig 1.a) increases linearly and the rate (Fig 1.b) stabilizes over time. It is also observed that an increment in the number of nodes or flows causes the rate to increase. This happens because more control packets will be created and eventually, the increment in the sum will be sharper. The abbreviations below are RW: Random Waypoint, RD: Random Direction , MG: Manhattan Grid.



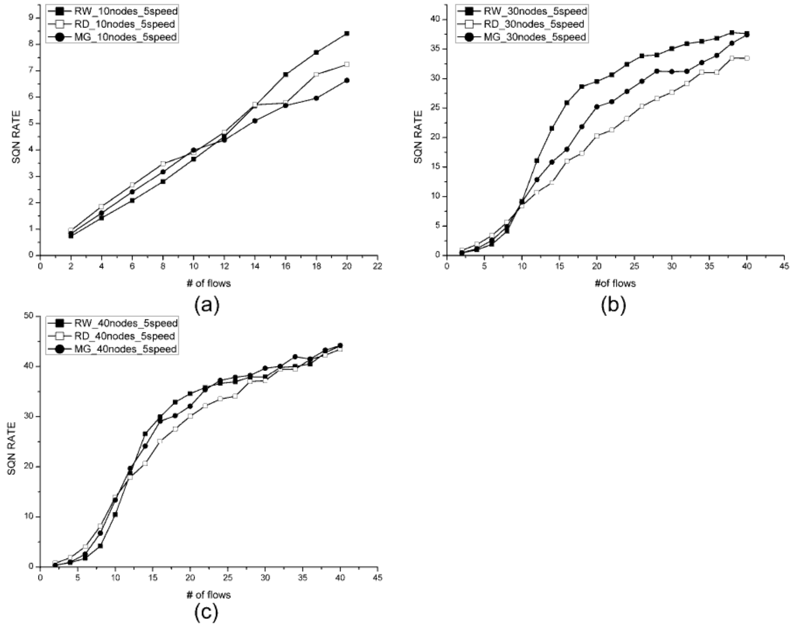
(a)

(b)

Figure 5: (a) The sum of SQNs of every node over time (b) The SQN rate of the network over time

The next step is to find out how the $SQN_{network\_rate}$ behaves according to the chosen variables. We define the $SQN_{network\_rate}$ as a function $SQN_{network\_rate}$=F(mobility model, number of nodes, maximum speed, number of flows). In the below figures the impact of the increment of the number of nodes the selected rate is shown. To begin with, it is seen that the number of nodes affects the gradient of the graph. For a network of 10 nodes, the sqn rate follows approximately a linear increment. For a network of 30 and 40 nodes the sqn rate starts to increase exponentially and as the number of flows increases the sqn rate continues to follow a linear increment. It is notable that in different mobility models the exponential and the linear gradient is different. In order to understand why this happens, an analysis on the mobility patterns needs to be done. In the Random Waypoint model, the nodes tend to move close to the center of the simulation area whereas in the Random Direction model, the nodes move in a straight line till the boundaries of the simulation area and afterwards, they change their direction and speed. In the Manhattan Grid, the nodes move on horizontal and vertical lines, that represent bidirectional roads of a city. The nodes have higher chance to continue straight ahead rather than turning when they reach a crossroad. For the last two from the above mobility models, the nodes have a considerable possibility of moving for a long time close to the boundaries of the simulation area which makes the network topology to change fast. Consequently we can claim that in the Random Waypoint a node is more likely to have more neighbors than in the other two mobility morels. As a result, the more neighbors a node has, more control packets will be created and the $SQN_{network\_rate}$ will increase faster. With this argument, the exponential gradient of figures 2(b) & 2(c) can be justified.

Moreover we have to think what happens to the creation of new control packets with the increment of flows. Suppose an Ad hoc network of N nodes. When the number of flows is low it is more likely to have a different pair of source and destination nodes for each network flow. Upon the increment of the number of flows the possibility of multiple network flows between an already existing pair of source an destination nodes increases. As a result, when the source node want to search the destination node, it creates one control packet rather than one control packet for each network flow and consequently the $SQN_{network\_rate}$ starts to decrease after a certain number for a network of N nodes. This argument explains the why the sqn rate in figures 2(b) & 2(c) increases linearly after a certain number of flows. This argument also explains why in figure 2(a) the increment follows a linear behavior. In figure 2(a) the number of nodes is 10 and as a result the possible combinations of source and destination nodes are lesser than the networks of figures 2(b) & (c) in which the networks consist of 30 and 40 nodes respectively.



(a)

(b)

(c)

Figure 6: SQN_rate/flows as the mobility model change for (a) 10, (b)30 and (c) 40 nodes

Another important issue which needs to be discussed is the impact of the speed of the nodes. The below figure presents the difference of the $SQN_{network\_rate}$ behavior, in all the mobility models, of networks with max speed of 5 and 20 m/s. It is observed that in the Random Direction and the Manhattan Grid mobility models the speed does not affect the significantly the $SQN_{network\_rate}$ since its' value is nearly the same. In the Random Waypoint now it is seen that the speed affects the SQN rate in a predictable way. The above observations, point out that the $SQN_{network\_rate}$ is a metric that is not affected significantly from the mobility of the network.

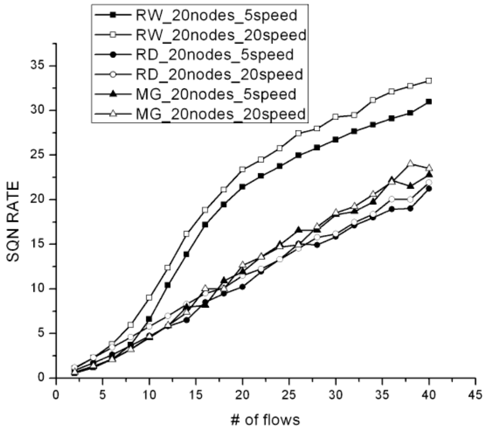


Figure 7: Here we see how the Speed increase affects the SQN rate behavior

Motivated by the fact that the rate of sum of the sequence numbers of a network is stable the next significant observation is the how the rate of each individual SQN behaves. The SQN of every node (defined as SQN_{node}) also increases linearly over time and consequently its' rate is also stabilized. Every node is responsible of increasing its' own SQN_{node} upon the creation of a RREQ/RREP message. Nodes that are either sources or destinations increase their sequence number faster.Consequently, their SQN_{node_rate} is stabilized in a higher value than the rate of the other nodes. In the below figure the SQN_{node_rate} rate of every node is shown for an ad hoc network of 20 nodes, moving with a speed of 10 m/s and the network flows are 10.

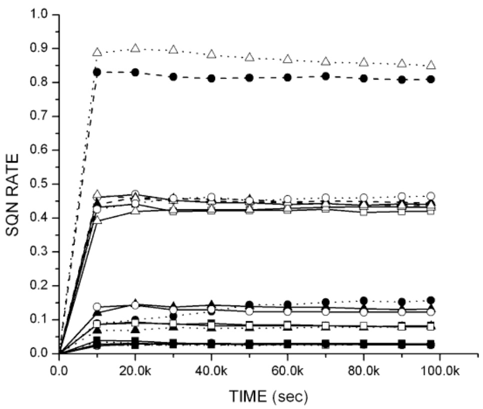


Figure 8: : The SQNnode rate for every node in the network of 20 nodes over time

Knowing that the SQN_{node_rate} of every node is stabilized, the only issue that needed to be verified is the behavior of the rate of the sum of the sequence numbers that are in the routing table of every node. In order to do that each time a node created a newly control packet, we calculated the rate of the sum of the sequence numbers in its' routing table plus its' own sequence number divided by the simulation time (defined as SQN_{total_rate}). This choice was done because the result will be closer to the $SQN_{network\_rate}$ compared to each individual SQN_{node_rate}. Moreover, it is also easier to be calculated by every node compared to the $SQN_{network\_rate}$ which requires knowledge of all the nodes in then network.Assuming that a node has N entries in its' routing table, $SQN_{tabnle-i}$ is the SQN in the i_{th} line, and K is the node's current sequence number, the SQN_{total_rate} at time T can be calculated as $[(\sum SQN_{tabnle-i})+K]/T$ (1). In all mobility models the mentioned rate also stabilizes over time. This happens because due to mobility issues, each node will have a sequence number for all the nodes in the network. In the below figure the SQN_{total_rate} is calculated for every node for a network of 20 nodes in the Random Direction model with the nodes moving with a speed of 10 m/s and the number of network flows are 10.

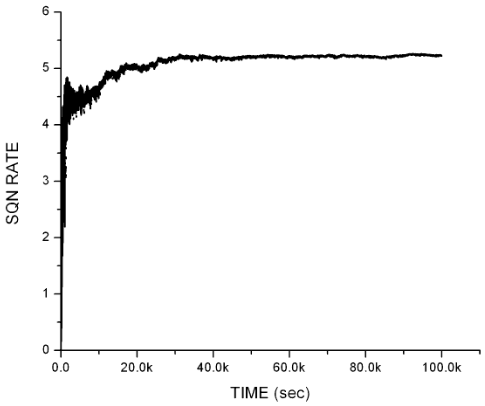


Figure 9: The rate of the sum of the sequence numbers in the routing table of every node plus its' own over time.

4. Blackhole Behavior

4.1 SQN behavior under blackhole attack

The black hole attack is a routing disruption attack in which the malicious node tries to act as a sinkhole for all packets. In order for this to be done, the malicious node replies to a legitimate RREQ with a malicious RREP containing a high sequence number. The node that created the RREQ trusts the malicious sequence number and forwards all the data through the malicious node. However, the malicious node instead of forwarding the packets it discards them.

In our simulations, we simulated ad hoc networks with a malicious black hole node in order to observe the difference between the normal behavior explained in the previous section. Let's assume that a malicious black hole node replies with a fake sequence number. This fake sequence number will be stored in some legitimate nodes' routing table. The nodes that get infected with the malicious sequence number are the nodes that create a path between the source node and the malicious node. If the infected nodes compute the rate described by equation (1) before and after receiving the malicious sequence number, they will notice a sudden increase in the above mentioned rate. The nodes that received the fake sequence number will accept an incoming sequence number from the victim node only when the victim's sequence number is greater than the fake one. Consequently we can claim that the infected nodes do not operate normally until their rate decreases to the normal value. In the below figure there is an example of an adhoc network of 20 nodes. Each node calculates the SQN_{total_rate} (1). An attacker replies only once to a legitimate RREQ at time 20000. It is clearly seen that the SQN_{total_rate} of the nodes that received this fake sequence number increased sharply. Moreover it is seen that those nodes operate normally when their rate has fallen back to its' normal value.

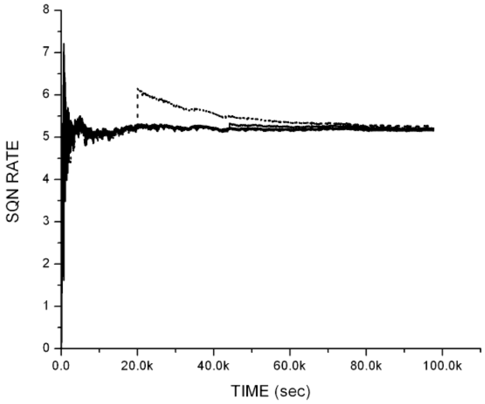


Figure 10: Malicious Behavior of the SQN

Furthermore, in our blackhole implementation the malicious node does not enter an arbitrary high sequence number. The malicious node is a part of the network and does not participate in the normal routing process. He overhears the sequence numbers that are exchanged between the legitimate nodes, stores them in his routing table, and when when he decides to attack, the malicious sequence number he sends the corresponding sequence number he knows for the victim node incremented by C. This is done in order for the source node to surely accept the fake RREP and consequently the malicious node. The legitimate nodes that receive the fake number will compute an increased rate calculated by equation (1). Moreover, The higher the fake sequence number is, the greater the calculated rate will be by the legitimate nodes. In the above figure (figure 6), the malicious node entered in the fake RREP the sequence number SQN_{victim}+10000 where SQN_{victim} is the SQN the malicious node knows about the victim node.

To sum up, the blckhole attack can have a significant impact on the normal routing operation of the network for a long time even if the malicious replies with a pretty low sequence number.Up to now, we studied the behavior of the sequence number of the AODV protocol and the conclusion is that the SQN_{total_rate} (1) can be a computational light way to detect blackhole attacks.

5. Proposed IDS

In this section, we propose an anomaly detection IDS based on the Cumulative Sum (CUSUM) algrotihm, which is one of the most commonly used algorithms based on the sequential change-point detection [19].The SQN_{total_rate} (1) depends on the size of the network, the mobility of the nodes, the mobility model and the ntework traffic. As a result it cannot be predicted and consequently we will use the nonparametric version of Recursive CUSUM [20][21][22] in order to build our detection mechanism. The recursive CUSUM begins with a random sequence X_n. This random sequence X_n has a mean value $\overline{X_n}$ and the values of this random sequence are close to the mean value $\overline{X_n}$ in normal operation, while in an abnormal operation, the values of the X_n are much higher than the mean value $\overline{X_n}$. However, the recursive CUSUM requires that the mean value $\overline{X_n}$ to be negative under normal operation[20]. Without any statistical loss, a new random sequence Z_n is defined such that $Z_n = X_n – C$ where $C \geq 0$. The mean value $\overline{Z_n}$ is negative and equals to $\overline{Z_n} = \overline{X_n} - C$. Finally the recursive CUSUM algorithm[20] can be used using the following equation: $Y_n = (Y_{n-1} + Z_n)^+$, $Y_0 = 0$, where x^+ equals to x if x > 0 and otherwise its' value is 0. The value of Y_n presents the cumulative positive values of Z_n and a large Y_n is a strong evidence of the occurrence of an attack. As a result a threshold can be defined and if Y_n becomes greater than the threshold value a change is detected.

In the below figure is observed how the three above mentioned random sequences behave. First of all, in top graph of figure 11, the Xn random sequence is seen. When a change takes place, it is seen that the value of the Xn is a+h. Consequently we can claim that we want to detect changes of minimum change of h. In the middle graph of figure 11, the Zn random sequence is seen. In this sequence we subtracted a positive value C such that the mean value of Zn before the change occurs is negative and close to zero. Upon a change the Zn becomes positive. Finally in the bottom graph of figure 11 the Yn random sequence is seen. Also a threshold N has been defined. When the Yn function becomes greater than the threshold N an attack is detected.

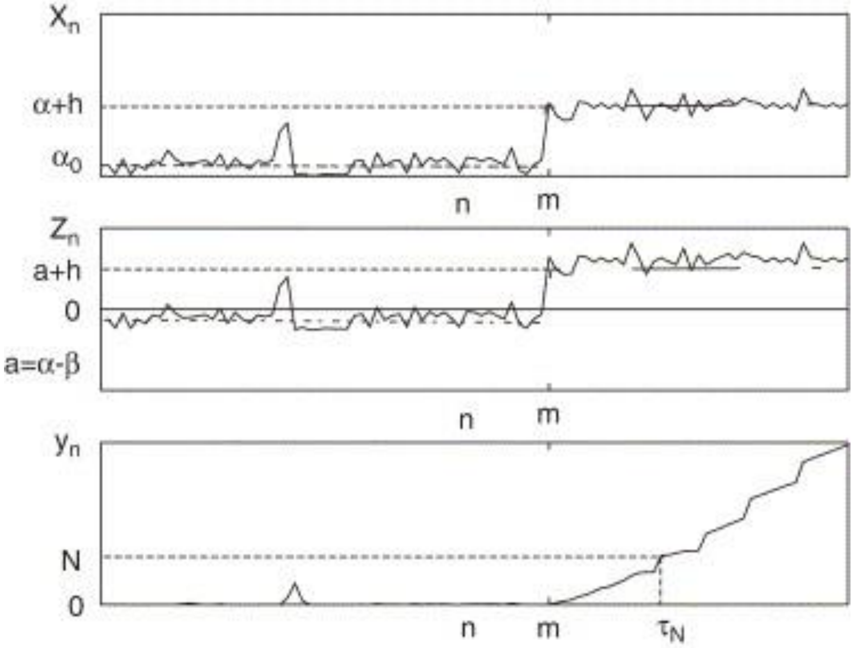


Figure 11: The random sequences Xn Zn Yn of the Recursive Cusum algorithm

At this point, the chosen random sequences for the ids should be explained. First of all, define F(t)=SQN_{total_rate}(t) (1). The first significant observation is that we could not use the F(t) as the definition of the X_n random sequence. This happens because the values of the F(t) can be positive or negative for a large period of time. This means that the variable C should be high enough in order for the CUSUM algorithm to work correctly. However we found out that in this way the ids could not detect attacks that have a small effect on the values of the X_n random sequence. As a result of the previous issue, define $X_n = F(n) - F(n-1)$ and X_0=0. This random sequence computes the difference between two sequential rates which are taken into certain time intervals. As time interval we used the value of 1 second. This value could have been higher but the variance of the values of X_n are greater as the time interval increases. In figure 7(a) the values of the random sequence Xn are shown. Also in this example there is a training phase of 10000 seconds in which the rate is being stabilized. It is clearly seen that at time 20000 an attack has occurred. At this point, the random sequence Z_n is being defined which has a negative mean value. This can be clearly seen in figure 7(b). It is notable that the figure

7(b) is a zoomed version and as a result the attack at time 20000 cannot be distinguished. Lastly in figure 7(c) the result of the Y_n function is shown. Also in that figure we define a threshold and it is clearly seen that the attack at time 20000 causes the Y_n to surpass the threshold value indicating the existence of an attack.

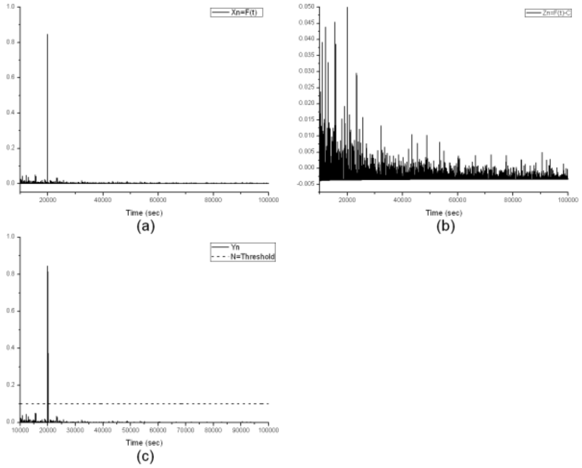


Figure 12: : (a) The sequence Xn=F(t) over time (b) The sequence Zn=Xn-C over time (c) The result of the cusum algorithm Yn, alongside with the detection threshold N

In the above figure the use of the cusum algorithm is shown in order to detect blackhole attacks. However in the above example, the attacker enters a high enough sequence number and consequently the attack is detected. The higher the malicious sequence number is sent, the higher the result of the Yn function will be. In order for the ids to run properly it should have the fewest possible false positives. This means that the threshold needs to be lowered as much as possible.

In the sequential change point detection, there are two very important performance metrics. The first one is the time needed for a successful detection and the second one is the average time between two consecutive false positive results. Consequently, we simulated networks without the existence of an attacker and we studied the number of false positive results as well as the average time between two consecutively false positive results for different threshold values and for all mobility models. In figure 8(a) and 8(c) it is observed how the numbers of false positive results behave with the increment of the threshold value while in figures 8(b) and 8(d) the average time between two consecutively false positive results is shown. In subfigures (a) and (b) the network simulated consists of 10 nodes and the network flows are 5 while in subfigures (c) and (d) the nodes are 20 and the flows are 10. In both simulations the nodes move with a max speed of 10 m/s. It is seen that as the threshold value increases the false positives decrease and consequently the average time between two consecutive false positive increases. The average time between the false positives is being calculated as the total time in which the IDS was running divided by the total number of false positives. As a result when the false positives are zero the average time between two false alarms cannot be calculated because the denominator of the fraction would be zero.

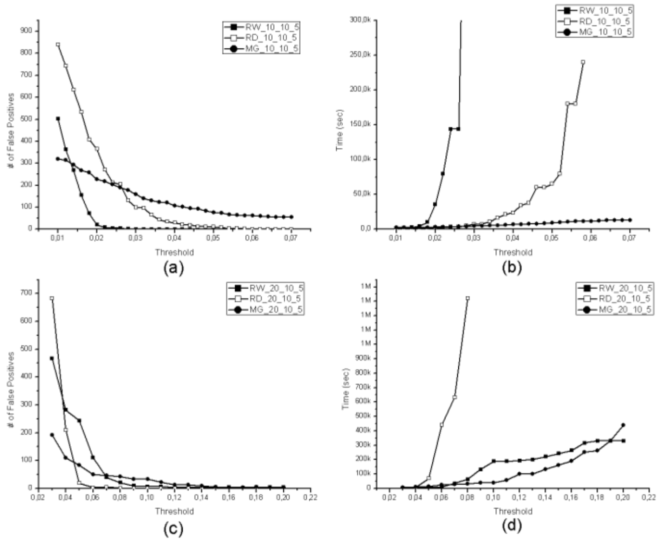


Figure 13: (a) Number of False Positives and (b) Average time between two consecutively false positives for networks of 10 nodes moving with 10 m/s and 5 flows. (c) Number of False Positives and (d) Average time between two consecutively false positives for networks of 20 nodes moving with 10 m/s and 5 flows

6. Conclusions and Future Work

MANETs is a recent field which has a wide range of applicability. Due to this fact, the MANETS need to be secured in order to prevent security attacks. In this paper we dealt with the Blackhole attack and the most important part of AODV, the sequence numbers. The first significant conclusion is that the sequence numbers behave in a similar way in big and small networks with different mobility models. Moreover, the $_{\text{SQNtotal\_rate}}(1)$ which is used as a metric is stabilized in networks of great mobility. As a result we used that metric in order to build an Intrusion Detection System based on the CUSUM algorithm. The first step of the evaluation of the proposed IDS has been done by examining the false positives and the average time between two consecutively false positives which both of them are two important metrics which apply to the sequential change point detection. The next step for the complete evaluation of the proposed IDS is the study of the false negative results and the effectiveness of the detection it offers.

7. References

[1] Rayal N. Raj and PrashantB.Swadas, *DPRAODV: A dynamic learning system against black hole attack in AODV based MANET,* International Journal of Computer Science Issues (IJCSI), Vol. 2, Issue 3, pp:54-59,2009

[2] Sanjay Ramaswamy, Huirong Fu, Manoharsreekantaradhya John Dixon and Kendall Nygard,*Prevention of Cooperative Black Hole Attack in Wirless Ad Hoc Networks,* 2003

[3] Mohammad Al-Shurman and Seong-Moo Yoon and Seungjin Park,Black Hole Attack in Mobile Ad Hoc networks

[4]Hongmei Deng, Wei li and Dharma P.Agrawal, *Routing Security in Wireless Ad Hoc Network,* IEEE Communications Magazine, Vol. 40, Issue 10, 2002

[5]Chang Wu Yu, Tung Kuang-Wu, ReiHeng, Cheng, abd Shun Chao Chang,*A Distributed and Cooperative black Hole Node Detection and Elimination Mechanism for Ad Hoc Networks*PAKDD 2007 Workshops, pp:538-549,2007

[6] B. Sun, K. Wu, Y. Xiao, R.Wang,*Integration of mobility and itrusion detection for wireless ad hoc networks,* Wieley International Journal of Communication Systems, Vol. 20,Issue 6, pp:695-721,2007

[7] B. Sun, K. Wu, UW. Pooch,*Routing anomaly detection in mobile ad hoc networks,* In Proc. of IEEE international conference on computer communications and networks (ICCCN'03),pp:25-31,2003

[8] B. Kannhavong, H. Nakayama, Y.Nemoto, N.kato, A. Jamalipour,*A Survey of Routing Attacks in mobile Ad Hoc networks,* In Wireless Communications, Vol. 14,Issue 5, pp:85-91.

[9] Fan Bai and Ahmed Helmy,*A Survey of Mobility Models in Wireless Adhoc Networks*

[10] Shiddhartha Raj Bhandari, GyuMyoung Lee, and Noel Crespi,*Mobility Model for User's Realistic Behaviour in Mobile Ad Hoc Network*

[11] P. Nain, D. Towsley, B. Liu, and Z. Liu,*Properties of Random Direction Model*, IEEE INFOCOM, Miami, FL, Mar. 2005

[12] E.M. Royer, P.M. Melliar-Smith, and L. E. Moser,*Analysis of the Optimum Node Density for Ad hoc Mobile Networks*, IEEE ICC, Helsinki, Finland, June 2001

[13] Johnson, David B.; Maltz, David A.,*5. Dynamic Source Routing in Ad Hoc Wireless Networks*, In Imieliński, Tomasz; Korth, Henry F..Mobile Computing. Springer,1996

[14]Satoshi Kurosawa, Hidehisa Nakayama, Nei Kato, Abbas Jamalipour and Yoshiaki Nemoto,*Detecting Blackhole Attack on AODV-based Mobile Ad Hoc Networks by Dynamic Learning Method*, International Journal of Network Security, Vol 5,No.3, PP. 338-346, Nov 2007

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

[16] Network Simulator ns-2 - http://www.isi.edu/nsnam/ns/

[17] Informatik 4: Bonnmotionhttp://net.cs.uni-bonn.de/wg/cs/applications/bonnmotion/

[18] EsaHyytiä, Pasi E. Lassila, Jorma T. Virtamo: Spatial Node Distribution of the Random Waypoint Mobility Model with Applications. IEEE Trans. Mob. Comput. 5(6): 680-694 (2006)

[19] M. Basseville and I.V. Nikiforov, Detection of Abrupt Changes: Theory and Application, Prentice Hall, 1993

[20]B.E Brodsky and B.S. Darkhovsky, Nonparametric Methods in Change-Point Problems, Kluwer Academic Publishers, 1993

[21] Rudolf B. Blazek, Hongjoong Kim, Boris Rozovskii, and Alexander Tartakovsky. A novel approach to detection of “denial-of-service” attacks via adaptive sequential and batch-sequential change-point detection methods. In Proceedings of IEEE Systems, Man and Cubernetics information Assurance Workshop, June 2001.

[22] Haining Wang, Danlu Zhang, and Kang G. Shin. Detecting SYN flooding attacks. In Proceedings of IEEE Infocom’2002, June 2002.