FAPRP: A Machine Learning Approach to Flooding Attacks Prevention Routing Protocol in Mobile Ad Hoc Networks

Ngoc T. Luong[1,2], Tu T. Vo[1], Doan Hoang[3]

[1] *Faculty of Information Technology, Hue University of Sciences, Hue University, Hue 530000, Vietnam.*

[2] *Faculty of Mathematics and Informatics Teacher Education, Dong Thap University, Dong Thap 870000, Vietnam.*

[3] *Faculty of Engineering and Information Technology, the University of Technology Sydney, Sydney 2007, Australia.*

Correspondence should be addressed to Doan Hoang; Doan.Hoang@uts.edu.au

Abstract. Request route flooding attack is one of the main challenges in the security of Mobile Ad hoc Networks (MANETs) as it is easy to initiate and difficult to prevent. A malicious node can launch an attack simply by sending an excessively high number route request (RREQ) packets or useless data packets to non-existent destinations. As a result, the network is rendered useless as all its resources are used up to serve this storm of RREQ packets and hence unable to perform its nomal routing duty. Most existing research efforts on detecting such a flooding attack use the number of RREQs originated by node per unit time as the threshold to classify attackers. These algorithms work to some extent; however, they suffer high misdetection rate and reduce performance of the network. This paper proposes a new flooding attacks detection algorithm (FADA) for MANETs based on a machine learning approach. The algorithm relies on the route discovery history information of each node to capture similar characteristics and behaviors of nodes belonging to the same class to decide if a node is malicious. The paper also proposes a new flooding attacks prevention routing protocol (FAPRP) by extending the original AODV protocol and integrating FADA algorithm. The performance of the proposed solution is evaluated in terms of successful attack detection ratio, packet delivery ratio, and routing load in both normal and under RREQ attack scenarios using NS2 simulation. The simulation results show that the proposed FAPRP can detected over 99% of RREQ flooding attacks and performs better in terms of packet delivery ratio and, and routing load compared to existing solutions for RREQ flooding attacks.

Keywords: AODV, FADA, FAPRP, MANETs, flooding attacks.

1. INTRODUCTION

A Mobile Ad hoc Network (MANET) [1] is a collection of wireless mobile devices (called nodes) that dynamically form a network in environments, such as disaster rescue, urgent conference or military mission, without the support of a network infrastructure. The topology of the network may change frequently because nodes can join or leave the network at will. In a MANET, nodes coordinate among themselves to maintain the connections among them. Data transfer from a source node to a non-neighbor destination node is routed through mediate nodes. A node can act as a host and a router at the same time. A network routing protocol in a MANET specifies how nodes in the network communicate with each other. It enables the nodes to discover and maintain the routes between any two of them. Many routing protocols have been developed for MANETs such as ad hoc on-demand distance vector (AODV) [2], dynamic destination sequenced distance vector (DSDV) [3], and zone routing protocol (ZRP) [4]. They are classified into three groups: proactive, reactive, and hybrid routing protocols. With proactive routing protocols, the routes between nodes need to be established before data packets can be sent. These protocols are suitable for fixed topology networks. In contrary, reactive routing protocols are suitable for dynamic topology networks as nodes only try to discover routes on demand. In complex network topologies, hybrid routing protocols are often used [5]. MANETs are thus essential in infrastructureless situations for communication, however, they suffer from various types of Denial of Service (DoS) attacks that deny user of a service or a resource he/she would normally expect to receive. Disrupting the routing services at the network layer is an example of DoS [6][7] where a malicious node (MN) tries to deplete resources of other nodes. Other types of DoS include Blackhole [8], Sinkhole [9], Grayhole [10], Whirlwind [11], Wormhole [12] and Flooding attacks [13]. Flooding attack is a particular form of DoS attacks in MANETs where malicious nodes mimic legitimate nodes in all aspects except that they do route discoveries much more frequently with the purpose of exhausting the processing resources of other nodes. This type of attacks is simple perform with on-demand routing protocols, typically as AODV [14]. Amongst HELLO, RREQ and DATA flooding attacks, route request (RREQ) flooding attacks is the most hazardous because it is easy to create a storm of request route packets and cause widespread damages. This paper focuses on the request route flooding attack.

Previous researches on RREQ flooding attacks mainly focus on detection algorithms that rely on sending frequency of RREQ packets [15-20]. Every node uses a fixed (or dynamic) threshold value to detect an attack. The threshold is calculated based on the number of RREQs originated by node per unit time. A node labels a neighbor node malicious if it receives a higher number of RREQs than the allowed threshold from the neighbor. These algorithms, however, have many weaknesses in dealing with the dynamics of MANETs. These include: (1) An algorithm with a fixed

threshold is not flexible and is not able to cope with dynamic environments where optimal threshold values vary accordingly; (2) Even with dynamic threshold algorithms, where the threshold takes into account other factors such as network traffic, mobility speed, and frequency of malicious node attacks, misclassifications rates are still high. In high mobility environments, the connection state of network nodes changes very frequently, a node may not be able to capture accurate and adequate information to distill it to a single threshold ; (3) A normal node may be mistaken for a malicious node even if it legitimately sends out a high number of route requests in response to a high priority event; or (4) A malicious node may avoid the threshold detection mechanism simply by sending RREQ packets at a frequency just lower the threshold value.

In this paper, we propose and investigate a different approach for detecting flooding attacks. Our solution relies on the route discovery history information of each node to classify a node as malicious or normal. The route discovery history of each node is represented by a of route discovery frequency vector (RDFV). The route discovery histories reveal similar characteristics and behaviors of nodes belonging to the same class. This feature is exploited to differentiate abnormal behavior from a normal one. RDFV is defined as the feature vector for detecting malicious nodes in MANET environment. We propose a flooding attack detection algorithm (FADA) to detect malicious node based on RDFV. We propose a novel flooding attacks prevention routing protocol (FAPRP) by incorporating the FADA algorithm and extending the AODV protocol. We evaluate the performance of our solution in terms of successful detection ratio, packet delivery ration, routing load in both normal and under RREQ attack scenarios using NS2 simulation. The simulation results showed that our approach can detected over 99% of RREQ flooding attacks, had better packet delivery ratio and, and routing load compared to existing solutions for RREQ flooding attacks, and introduced negligible overhead relative to AODV for normal scenarios. In this paper, the main contributions are as follows:

(1) Introduced a new route discovery history measure, the vector of route discovery frequency (RDFV), to capture the behavior of MANET nodes.
(2) Proposed a flooding attack detection algorithm (FADA), a k-nearest neighbors-based machine learning algorithm, using RDFV dataset to detect malicious nodes.
(3) Proposed a flooding attack prevention routing protocol (FAPRP) by integrating FADA into the original AODV protocol.
(4) Evaluated the effectiveness and the performance of the proposed solution for high-speed mobility MANETs under RREQ flooding attacks.

The remainder of this paper is structured as follows: Section 2 presents a review of the related work on detection of flooding attacks. Section 3 presents our solution and a novel flooding attacks prevention routing protocol (FAPRP) by improving

AODV protocol using FADA. Section 4 presents the results of evaluation the performance of the proposed solution relative to existing solutions. Section 5 concludes the paper.

2. RELATED WORKS

2.1. Overview of AODV

AODV is a popular reactive routing protocol in which a node only initiates the process for finding a path to the destination if it wants to send data. Basically, when the source node (N_S) wants to communicate with the destination node (N_D), without an already discovered route to the destination, N_S starts a route discovery process by broadcasting a route request (RREQ) packet containing the destination address. The nodes that receive the packet will in turn broadcast it. When N_D receives the packet, it will send a route reply (RREP) packet back to source node. Once a route has been discovered, HELLO and RERR packets can be used to maintain the status of the route.

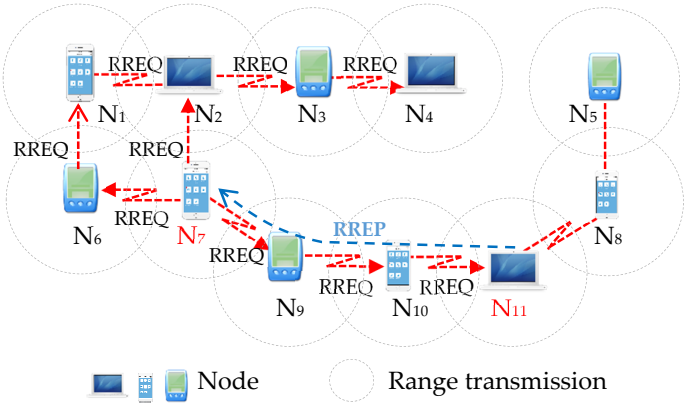


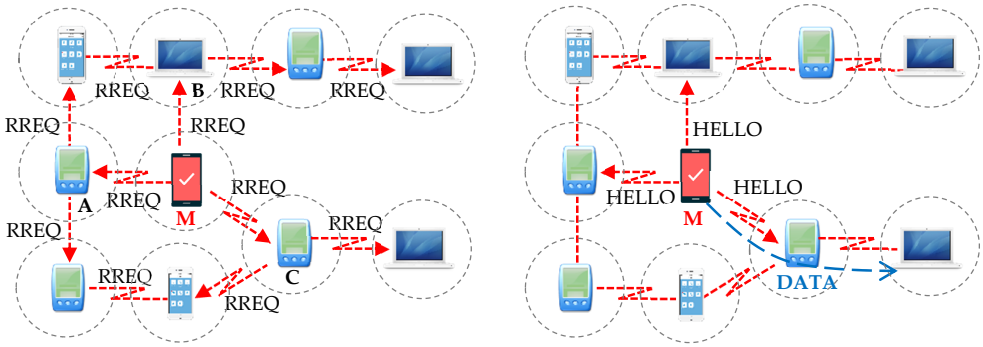
Figure 1. Description of route discovery process of AODV in the MANET

Figure 1 describes the route discovery process of AODV, source node (N_7) discovers route to destination node (N_{11}) by broadcasting an RREQ to its neighbor nodes. When a node receives the RREQ packet for the first time, it broadcasts the packet and sets up a reverse path to the source. If the node receives the same RREQ subsequently, it simply drops the packet. When N_{11} gets a RREQ, it unicasts a RREP packet to the source node through the established reverse {$N_{11} \rightarrow N_{10} \rightarrow N_9 \rightarrow N_7$}. When N_7 gets a RREP, it establishes successfully a new path to N_{11} with 3 hops routing cost and adds the new entry into its routing table.

2.2. Flooding attacks on AODV

Flooding attack is a form of DoS attacks in which malicious nodes broadcast the false packet in the network to exhaust the resources and disrupt the network operation. Depending upon the type of packet used to flooding the network, flooding attack can be categorized in three categories, RREQ, DATA and HELLO flooding

attack. In RREQ flooding attack, a malicious node continuously and excessively broadcasts fake RREQ packets, which causes a broadcast storm and floods. The RREQ flooding attack is considered most harmful in MANET because it can ruin the route discovery process by exhausting the channel bandwidths and the processing resources of affected nodes. In DATA flooding attack, a malicious node can excessively broadcast data packets to any nodes in the network. This type of attacks has more impact on the nodes participating in the data routing to the destinations. In HELLO flooding attack, nodes periodically broadcast HELLO packets to announce their existence to their neighbors. A malicious node abuses this feature to broadcast HELLO packets excessively and forces its neighbors to spend their resources on processing unnecessary packets. This type is only detrimental to the neighbors of a malicious node. Figure 2 shows the behavior of malicious nodes (M) in a MANET for these types of attacks.



a) RREQ flooding b) HELLO and DATA flooding

Figure 2. Description of flooding attacks in the MANET

2.3. Review on related research

This section summarizes related work on threshold-based, machine learning-based, hash function-based and digital-signature-based approaches in detecting and preventing flooding attacks in MANETs. Table 1 summarizes of these methods and their drawbacks.

2.3.1. On fixed threshold-based approach

Solutions are simple with a fixed threshold for mitigating the impact of RREQ flooding attacks. However, with static threshold, these methods are not suitable for dynamic environments where nodes are highly mobile and frequently broadcast route request packets. In [15], Gada used three fixed thresholds: RREQ_ACCEPT_LIMIT, RREQ_BLACKLIST_LIMIT and RATE_RATELIMIT. The default value of RATE-RATELIMIT is 10. If the rate of receiving request packets is greater than RREQ_ACCEPT_LIMIT but less than RREQ_BLACKLIST_LIMIT, packets are simply dropped and not processed. If it is greater than RREQ_BLACKLIST_LIMIT, the source is declared as a malicious node. The weakness

of this solution is that also it may lead to blacklisting of normal nodes false positive problem [16] and causes excessive end-to-end delay by dropping legitimate request packets once the RREQ_ACCEPT_LIMIT threshold is crossed.

In ([13][17]), the authors developed Flooding Attack Prevention (FAP) that prevents RREQ and DATA flooding attacks in MANETs. They argued that the priority of a node is adversely proportional to its broadcast frequency of RREQ. Hence, nodes that generate a high frequency of route requests will have a low priority and may be removed out of the routing process. It is suggested that a node should not originate more than 10 RREQ packets per second, and hence, the threshold of FAP is set at 15.

In [16], Song proposed a simple technique using an Effective Filtering Scheme (EFS) to detect malicious nodes. This solution uses two limit values: RATE_LIMIT and BLACKLIST_LIMIT. If the detected RREQ rate is higher than the RATE_LIMIT and the BLACKLIST_LIMIT, the malicious node is declared and it will be put into the black list. If the rate of RREQs originated by a node is between the RATE_LIMIT and the BLACKLIST_LIMIT, the RREQ packet is added to a "delay queue" waiting to be processed. Here authors set the RATE_LIMIT threshold to 5 and set the BLACKLIST_LIMIT up to 10.

2.3.2. On dynamic threshold-based approach

Solutions with dynamic thresholds are more flexible as they can cope with the dynamic environment of MANETs. In [18], Mohammad proposed an improved protocol called B-AODV. In this method, each node employs a balance index (BI) for acceptation or rejection of RREQ packets. If the RREQ rate is higher than the BI value, a malicious node is defined and the RREQ packet is dropped. The results showed that B-AODV is resilience against RREQ flooding attacks. The main drawback of B-AODV is that it may drop legitimate request packets of the node moving at high speed as the number of request packet may be higher than the balance index value [19]. Also, the method does not have a confirmation mechanism which can identify the node properly as a malicious node.

In [19], Gurung proposed a new mechanism called Mitigating Flooding Attack Mechanism. The mechanism is based on a dynamic threshold and consists of three phases. It deploys special Flooding Intrusion Detection System (F-IDS) nodes to detect and prevent flooding attack. The F-IDS nodes are set in the promiscuous mode to monitor the behaviour of nodes in the network. The proposed mechanism has several features: (1) it uses a dynamic threshold; (2) it has a confirmation mechanism in which the special F-IDS node confirms the node as a malicious node by sending a dummy reply packet and waits for the data packets; and (3) it has a recovery mechanism that allows the node to participate in the network after the expiry of the blocking time period. However, the use of several F-IDS nodes to monitor their

neighbours and to communicate among them limits the performance of the overall network, especially when the network is not under attack.

In [20], Tu introduced security mobile agents (SMA) to detect flooding attacks. An improved protocol, SMA2AODV, is proposed by integrating these SMAs into the discovery route process of the AODV protocol. During the training period, SMA agents are used to collect information for determining the minimal time-slot (the minimum time-slot for successfully discovered a path from a source node to a destination node) of the system (TS_{min}). After the training phase, all node (N_i) checks the security of the RREQ packet received from source node N_j before broadcasting it to the neighbors. If route discovery time-slot is smaller than the minimal time-slot of the system ($T < TS_{min}$), a Flooding attack is said to have occurred with N_j as the attacker. N_i then adds N_j into its black list. All RREQ packets of nodes in the black list will be dropped. The drawback of this method is that TS_{min} is only valid if no malicious node exist during the trainig period.

2.3.3. On machine learning approach

In [21], Patel proposed the use of Support Vector Machine (SVM) algorithm for detecting and preventing flooding attacks . The behavior of every node is collected and passes to the support vector machine to decide if a node is malicious based on a threshold limit.

In [22], Wenchao proposed a new intrusion detection system based on k-nearest neighbors (kNN) classification algorithm in wireless sensor network to separate abnormal nodes from normal nodes by observing their behaviors. An m-dimensional vector is used to represent nodes and their behavious such as the number of routing messages that can be sent in a period of time, the number of nodes with different destinations in the sending routing packets, the number of nodes with the same source node in the receiving routing packets. The test results show that the system has high detection accuracy. The paper, however, does not present the algorithm for building training data sets.

Table 1. Summary of drawbacks of related works for detecting flooding attacks

Ref	Name	Year	Method	Drawback
[15]	Proposed-AODV	2004	Fixed threshold	Uses static threshold value which is not suitable for high mobility environment. Malicious node can pass the security mechanism by transmitting the RREQ packet at a frequency lower than the threshold.
[13]	FAP	2005		
[16]	EFS	2006		

[18]	B-AODV	2016	Dynamic threshold	It can drop the valid request packet of the node moving with high mobility speed if the number of request packet is greater than BI value. Malicious node can pass the security mechanism by transmitting the RREQ packet at a frequency lower than the threshold.
[19]	F-IDS	2017		Performance depend on some assumptions. Using new control packets (ALERT) will increase communication overhead and limination performance when operating in network environment without attacks. Malicious node can pass the security mechanism by transmitting the RREQ packet at a frequency lower than the threshold.
[20]	SMA$_2$AODV	2017		There are not any malicious node exist in scenario during threshold value make phase. Malicious node can pass the security mechanism by transmitting the RREQ packet at a frequency lower than the threshold.
[21]	SVMT	2013	SVM	Proposed algorithm uses fixed threshold to detect malicious nodes.
[22]	kNN-AODV	2014	kNN	Accordance with the requirement of wireless sensor network intrusion detection. The algorithm for building training data sets, used in the kNN algorithm has not been clearly presented.

3. THE PROPOSED FAPRP SOLUTION

This section we present our algorithms and routing protocol for detecting flooding attacks in MANETs. First, we define the feature vector that represents the behavior of a node based on its history of rout discovery: the route discovery frequency vector (RDFV). Second, we describe an algorithm for obtaining the

training dataset which describes the normal behavior and the abnormal behavior of nodes for normal/malicious classification. Third, we present our flooding attack detection algorithm, and finally we present our proposed AODV-based flooding attacks prevention routing protocol. Table 2 defines symbols used in the paper.

Table 2. Description of symbols

Variable	Description
t_i	Route discovery time i^{th}
T_i	Route discovery time slot i^{th}
V_{Ns}	Vector of route discovery frequency of Ns node
n	Size of vector of route discovery frequency
k	Cutoff value for kNN algorithm

3.1. Route discovery frequency vector

In order to detect RREQ flooding attacks with kNN, the crucial problem is the selection of a feature vector that maximizes the separation of the normal and the malicious data classes and produces highly reliable classification. The selected features should be able to succinctly capture the inherent behavior of a node performing RREQ requests and the time-related network activities through their historical data records in order to differentiate "normal" from and "malicious" behavior. We propose a route discovery frequency vector as the feature vector for this purpose. To quantify this vector we define the following terms.

- **Definition 1:** *Route discovery time (t_i),* is the duration from the time a node first broadcasts a route discovery packet to the time it receives the corresponding route response. Assuming that node N_i receives the i^{th} RREQ packet from the source node Ns at time s_i and N_i receives the route response packet at time e_i, the route discovery time (t_i) is defined by eqn 1.

$$t_i = e_i - s_i \tag{1}$$

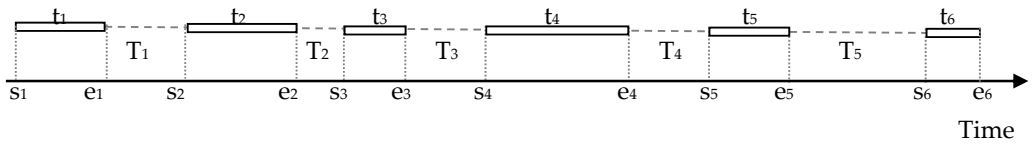
.

- **Definition 2:** *Inter-route discovery time (T_i),* is the duration from the end of a route discovery to the beginning of the next route discovery. Assuming that the node N_i receives the $i+1^{th}$ RREQ packet from the source node Ns at time s_{i+1}, the inter-route discovery time (T_i) is defined by eqn 2.

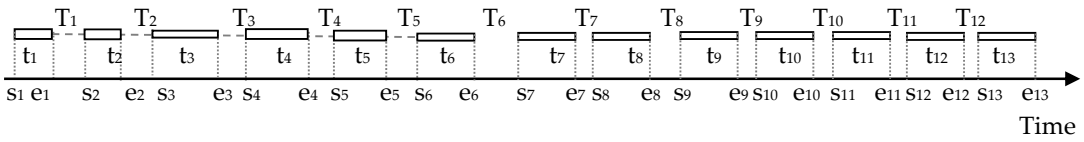
$$T_i = s_{i+1} - e_i \tag{2}$$

In AODV routing protocol, route discovery frequency of a node depends on how frequent the node has to find a path to carry data to its destination. All normal nodes have route discovery frequencies within a range, but malicious nodes have higher route discovery frequencies as their aim is to flood the network. Consider Figure 2a,

it shows three normal nodes A, B, C and one malicious node M. Figure 3a shows the route discovery history of the normal node (C) as recorded by the normal node (A). Figure 3b shows route discovery history of the malicious node (M) that it is also recorded by the normal node (A). The figures show that node C sent 6 RREQ packets and node M sent 13 RREQ packets over roughly the same duration.



a) Route discovery history off normal node (C)



b) Route discovery history of malicious node (M)

Figure 3. Route discovery history recorded at normal node (A)

We use a *n*-dimensional vector V_{Ni} (a_1, a_2, a_3, …a_n) to represent route discovery history of node N_i, where n is the size of the vector, and a_i is the i^{th} inter-route discovery time.

Example: Route discovery history of the malicious node shown in Figure 3b is represented by vector $V_M(T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9, T_{10}, T_{11}, T_{12})$ of size 12.

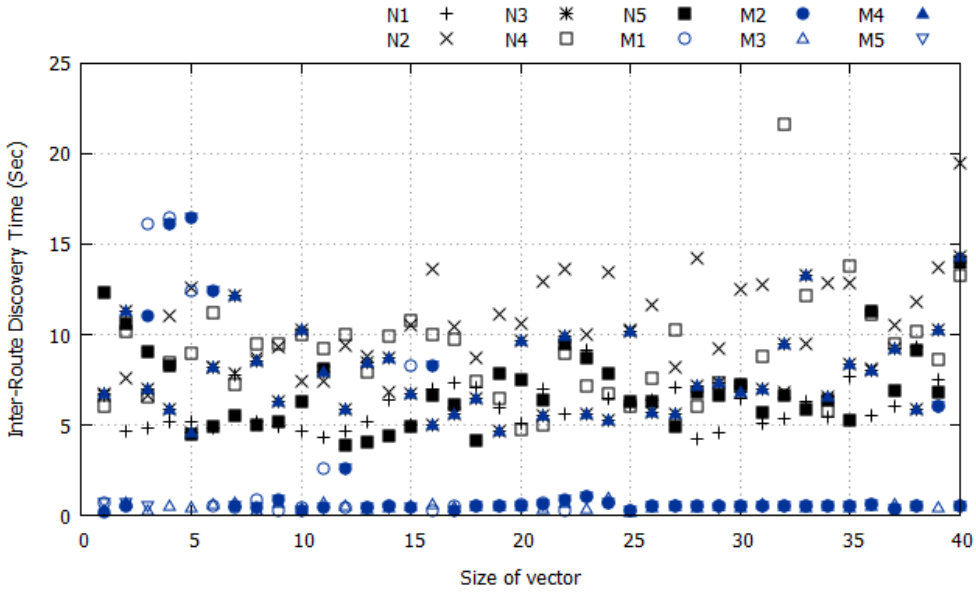


Figure 4. An example describes vectors of route discovery frequency of 5 normal nodes (N_1 to N_5) and 5 malicious nodes (M_1 to M_5).

Figure 4 shows typical vectors of size 40 of the route discovery frequency of normal and malicious nodes, by NS2 simulation. It can be seen that the inter-route discovery time values for all normal nodes (N_1 to N_5) are generally larger (> 1 sec) than those for malicious nodes (M_1 to M_5) as they have low route discovery frequencies. However, there are cases where the malicious inter-route discovery times (Ts) are indistinguishable from the normal ones. One reason for this is the mobility of nodes in the environment; on moving, a recording node may not receive RREQ packets from a malicious node until some later time. Another reason for is that a RREQ may be delayed in a waiting queue before being relayed, resulting in a larger value of T. Other reason for the overlapping region is when a malicious node floods the network at a frequency close to the rate at which a normal node can generate RREQs. As demonstrated in section 4, our proposed algorithm successfully recognizes these abnormal cases based on route discovery frequency feature.

3.2. Algorithm for obtaining a training dataset

We use NS2 [23] – version 2.35 to build a training dataset of NVC and MVC vector classes. The simulation scenario is set up with 100 normal nodes and 1 malicious node, operated in the area of 2000m x 2000m. Normal nodes move under random waypoint model with maximum speeds 0m/s, 10m/s, 20m/s, 30m/s and 40m/s scenarios; a malicious node is positioned at the center (1000m x 1000m) as shown in Figure 5. Other simulation parameters include: AODV routing protocol, 50 UDP connections, constant bit rate (CBR) traffic type, the first data source commences at time 0, other data sources commence at 5 seconds apart after the first, the malicious node respectively floods *f* packets every second (*f* may take on different values: 2, 5, 10, 50, and 100).

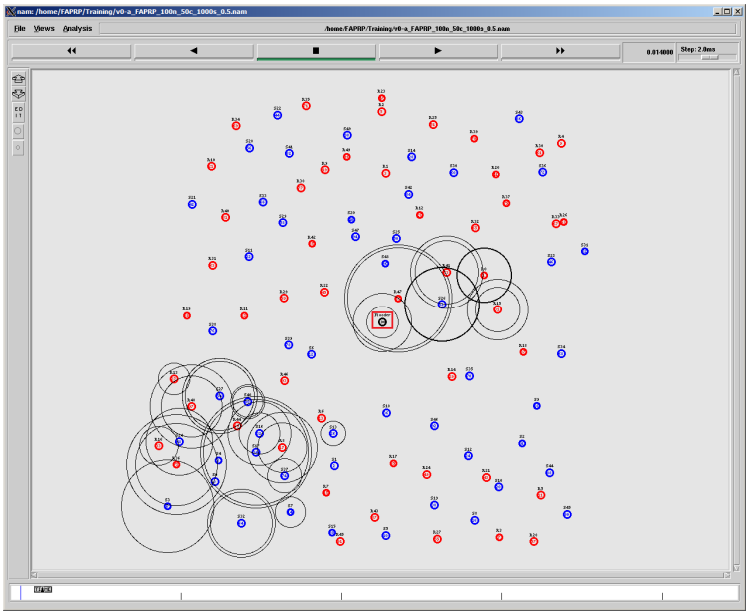


Figure 5. Static network topology simulation for training, 50 UDPs connections, malicious node positioned at the square in the center

The training process proceeds as follows.

Step 1: Select the dimension or size (n) of the feature vectors;

Step 2: Set the frequency of flooding to 2 initially ($f = 2$ per second);

Step 3: For each of the mobile scenarios (0m/s, 10m/s, 20m/s, 30m/s, and 40m/s), simulate the MANET as follows. Each node records the inter-route time of a source node (T_i) on receiving a RREQ from the source node i. Add T_i to the malicious history frequency vector if the source is malicious, otherwise it is added to the normal history frequency vector. At the end of this step for each scenario, two sets of vectors are established:

- 100 Malicious vectors: $V_M^j\left(T_1^M, T_2^M, T_3^M, ..., T_n^M\right); \forall j = \overline{1..100}$
- 100 Normal vectors: $\overline{V^j}\left(\frac{\sum_{i=1}^{50} T_1^i}{50}, \frac{\sum_{i=1}^{50} T_2^i}{50}, \frac{\sum_{i=1}^{50} T_3^i}{50}, ..., \frac{\sum_{i=1}^{50} T_n^i}{50}\right); \forall j = \overline{1..100}$

Step 4: At the end of step 3 for all 5 scenarios, 100 average vectors for MVC and 100 vectors for NVC are obtained for this particular flooding frequency (f=2);

Step 5: The algorithm continues to establish MVC vectors and NVC vectors for other flooding frequencies ($f = 5$, 10, 50 and 100).

As the result of the training process, a training dataset with MVC and NVC vectors are shown in Figure 6. The training data set is used to classify an unknown sample vector V (in the next section). In Figure 6, each vector is of size 60. It can be seen that there is an overlap between the two classes due to node mobility as well as the closeness of the rate of generation of RREQ packets of malicious and normal nodes.

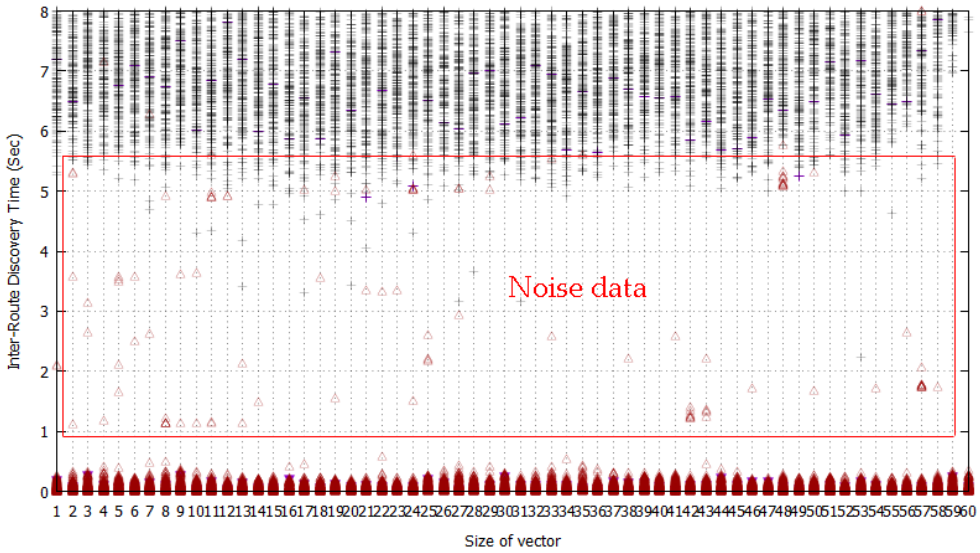


Figure 6. Two vectors class, black for NVC and red for MVC

3.3. Flooding attack detection algorithm (FADA)

All normal nodes collect route discovery information of source nodes in the network. On receiving a RREQ packet, a node employs the route discovery frequency vector (V_{Ns}) and uses a machine learning algorithm to determine if the source node is normal or malicious. The kNN-Classifier based on kNN [24] algorithm is utilized to classify the two classes based on the route discovery frequency vectors for NVC or MVC. The kNN algorithm is theoretically mature with low complexity that is widely used for data mining. The main idea is that if most of its k-nearest neighbor belong to a class, the sample belongs to the same class. In kNN, the nearest neighbor refers to the distance between two samples and various distance metrics can be used based on the feature vector that represents the samples. One of the most popular choices is the Euclidean in (3) to calculate the distance between V_1 and V_2. Algorithm 1 describes our algorithm for recognizing malicious nodes.

$$d(V_1, V_2) = \sqrt{\sum_{i=1}^{n} \left(V_1^i - V_2^i\right)^2} \quad (3)$$

Algorithm 1: Flooding attack detection algorithm using kNN

Input: Two class NVC and MVC, vector of route discovery frequency (V_{Ns})
Output: True if V_{Ns} in NVC, else return False

```
Begin
    MAX_VECTOR = 500;
    Double Array disMVC [MAX_VECTOR], disNVC [MAX_VECTOR];
    For int vt = 1 to MAX_VECTOR do {
        disMVC[vt] = Euclidean (VNs, MVC.Vectors[vt]);
        disNVC[vt] = Euclidean (VNs, NVC.Vectors[vt]);
    }
    Sort (disMVC and disNVC, ASC); // ascending sort
    int k1 = k2 = 0;
    While (k1 + k2 < k) {
        if (disNVC[k1] < disMVC[k2]) k1++;
        else k2++;
    }
    Return (k1 > k2);
End
```

3.4. FAPRP - A novel flooding attacks prevention routing protocol

In Mobile Ad hoc Network, a source node sends and receives packets through its neighbor nodes. If all neighbor nodes of the source node reject packets, it will be isolated and cannot communicate with the other nodes in its network. In Figure 2a,

the malicious node M broadcasts fake RREQ packets through nodes A, B and C. If the neighbor nodes A, B and C reject packets from M, node M cannot carry out its malicious behavior.

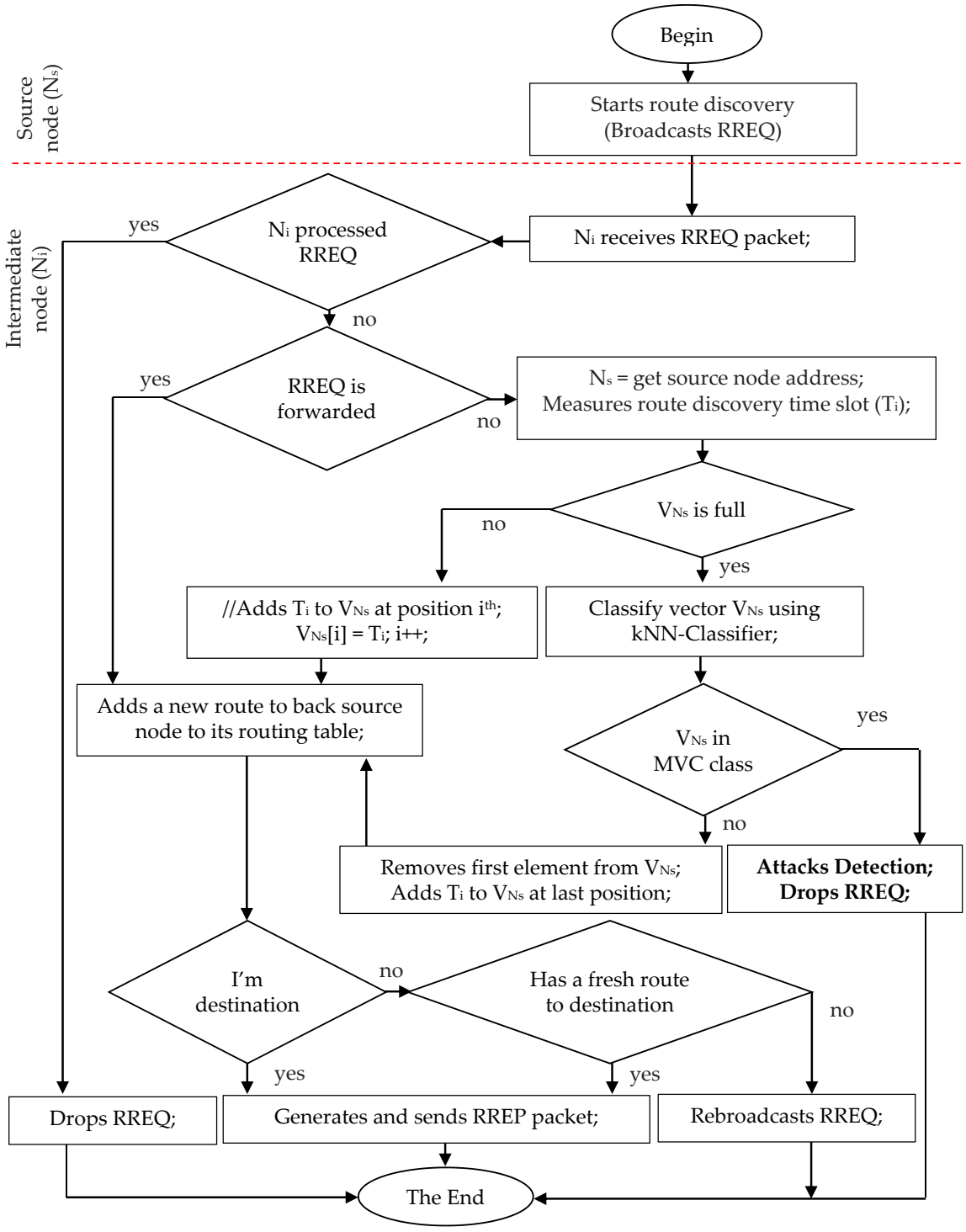


Figure 7. Request route process of FAPRP routing protocol

In the original AODV protocol, as intermediate nodes accept all RREQ route discovery packets from any source nodes, hackers may exploit this vulnerability to perform RREQ flooding attacks. We propose the flooding attacks prevention routing

protocol (FAPRP) by introducing the flooding attacks detection algorithm (FADA) into the route request phase of the AODV protocol. In FAPRP, only the source node's neighbor nodes use FADA to detect RREQ flooding attack on receiving RREQ packets. Other nodes forward RREQ packets without checking for RREQ flooding attacks.

Figure 7 describes how FAPRP detects an RREQ flooding attack when an intermediate node (N_i) receives an RREQ packet from the source node (N_s). When N_i receives an RREQ packet, if it is not a neighbor node of Ns, it broadcasts the RREQ packet without checking for RREQ flooding attacks; otherwise, it handles RREQ packet as follows:

- If the route discovery frequency vector of source node (V_{Ns}) is not full, N_i measures T_i and adds T_i to V_{Ns}, and then broadcasts the RREQ packet;
- Else, N_i uses FADA to classify Ns using its feature vector V_{Ns}.
  - If the source node is classified malicious, the RREQ packet is dropped and the algorithm terminates.
  - Else, N_i removes first element from V_{Ns} and adds T_i to the last position of V_{Ns}; and then broadcasts the RREQ packet;

4. PERFORMANCE EVALUATION BY SIMULATION

In this section, we use NS2 [23] – version 2.35 to evaluate the impact of RREQ flooding attacks on AODV and the proposed FAPRP protocol.

4.1. Simulation settings

Our simulation scenarios cover a 1000 meter by 1000 meter flat space, accommodating 50 normal mobile nodes. We consider 2 scenarios: one with a malicious positioned at the centre (Fig. 8a) and the other with two malicious nodes positioned as shown in Fig. 8b. Each malicious node may flood the network at the rate of 10 or 20 packets per second.

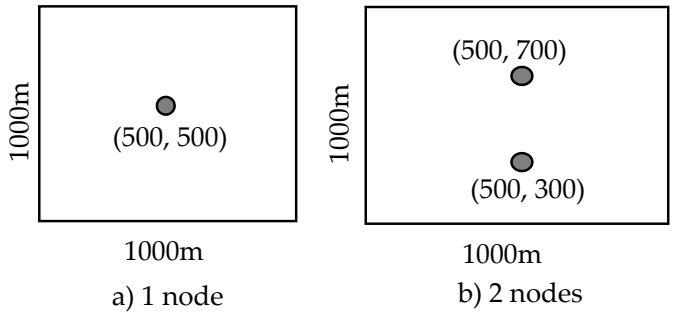


Figure 8. Malicious nodes location

The random waypoint [25] model is utilized as the mobility model. The minimum node speed for the simulations is 1 m/s while the maximum is 30 m/s. In

each simulation scenario, 20 sources transmit data at a constant bit rate (CBR). Each source transmits 512-byte data packets at the rate of 2 packets/second. The first source emits data at time 0, the following sources transmit data at 10 seconds apart. All parameters are described in Table 3.

Table 3. Simulation parameters

Parameters	**Setting**
Simulation area	1000 x 1000 (m^2)
Simulation time	500 (second)
Number of normal nodes	50 (nodes)
Node transmission range (R)	250 (m)
Number of malicious nodes	1, 2 (nodes)
Attacks frequency	10, 20 (packet/second)
Maximum speeds	1..10, 1..20 and 1..30 (m/s)
Traffic type	CBR (constant bit rate)
Transport protocol	UDP
Traffic type	CBR (constant bit rate)
Number of traffic	20
Data rate	2 (packet/second)
Packet size	512 (bytes)
Queue type	FIFO (DropTail)
Routing protocols	AODV, B-AODV [18], FAPRP
Size of vector (n)	10, 15, 20, 25, 30, 35, 40 and 60
Cutoff value (k)	10, 15, 20, 25, 30, 35, 40, 45 and 50
Distance type	Euclid

We evaluate the original AODV, the B-AODV, and our proposed protocol (FAPRP) and compare their performance with and without RREQ flooding attacks in terms of Attacks detection ratio, Packet delivery ratio, End-to-end delay, and Routing load metrics.

- Attacks detection ratio (ADR) is calculated using equation (4). AT is the number of RREQ packets that are accepted true, the packets come from normal nodes; AF is the number of RREQ packets that are accepted false, the packets come from malicious nodes; DT is the number of RREQ packets that are dropped true, the packets come from malicious nodes; DF is the number of RREQ packets that are dropped false, the packets come from normal nodes;

$$ADR = \frac{AT + DT}{AT + AF + DT + DF} * 100\,\% \tag{4}$$

- Packet delivery ratio (DPR): The ratio of the received packets by the destination nodes to the packets sent by the source nodes (eqn 5).

$$PDR = \frac{\sum_{i=1}^{n} DATA_i^{recieved}}{\sum_{j=1}^{m} DATA_j^{sent}} * 100\% \quad (5)$$

- End-to-end delay (ETE): This is the average delay between the sending time of a data packet by the CBR source and its reception at the corresponding CBR receiver (eqn 6), where T_{DATA}^{i} is the delay time for sending ith data packet to its destination successfully.

$$ETE = \frac{\sum_{i=1}^{n} T_{DATA}^{i}}{n} \quad (6)$$

- Routing load (RL): This is the ratio of the overhead control packets sent (or forwarded) to successfully deliver data packets. Routing discovery packets including: legitimate RREQ, fake RREQ, RREP, HELLO and RERR packets.

$$RL = \frac{\sum_{j=1}^{m} CONTROL\_PACKET_j^{overhead}}{\sum_{i=1}^{n} DATA_i^{recieved}} * 100\% \quad (7)$$

4.2. Simulation results

4.2.1. Effects of flooding attacks on the original AODV protocol

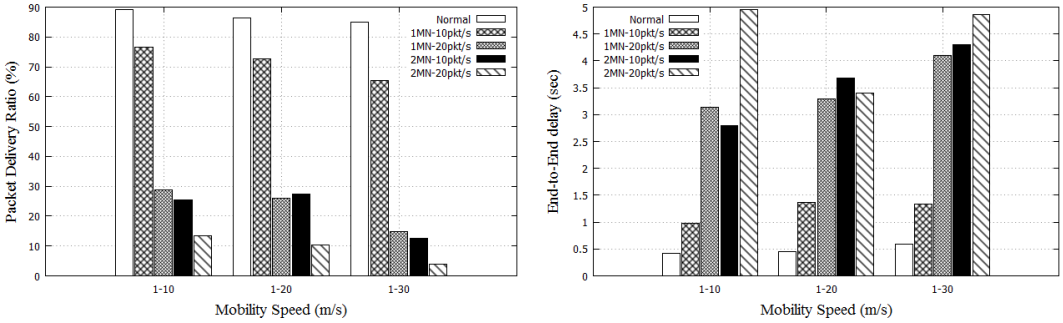
In this section we evaluate the performance of the AODV protocol with and without RREQ flooding attacks. We simulate 15 scenarios to evaluate the impact on the performance of AODV in terms of the above 4 defined metrics under various conditions including node mobility speeds, flooding frequencies, and malicious nodes. The main purpose of an RREQ flooding attack is to inject a large number of fake RREQ packets into the network making it less efficient in delivering legitimate packets. This effect is equivalent to handling excessive overhead packets causing a decrease in the network's packet delivery ratio, an increase in the average end-to-end packet delay, and an increase in the network's routing load. The simulation results are shown in Table 4.

Table 4. AODV performances under flooding attacks

MN	PDR (%)			RL (pkt)			ETE (sec)		
	1..10m/s	1..20m/s	1..30m/s	1..10m/s	1..20m/s	1..30m/s	1..10m/s	1..20m/s	1..30m/s
0	89.19	86.28	84.90	3.85	4.66	5.64	0.420	0.449	0.595
1	28.75	26.03	14.74	139.08	155.85	288.98	3.143	3.290	4.108
2	13.36	10.33	3.83	464.98	624.38	1,700.94	4.959	3.397	4.860

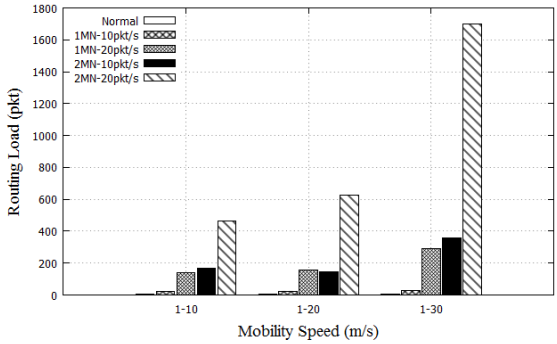
Figure 9 shows that the packet delivery ratio decreases, the routing load increases, and the end-to-end delay increases when the intruder floods attacking

packets. Figure 9a shows that without flooding attack, the AODV packet delivery ratio is above 84.9% and most packets reach their the destination nodes. However, the packet delivery ratio reduced drastically to 13.36% when the intruder uses 2 malicious nodes and floods 20 packets every second. Figure 9b shows the average end-to-end delay increases as the flooding attack frequency increases. When the attacker uses 1 malicious nodes and broadcasts 10 RREQ packets every second, the average end-to-end delay changes from 0.42s before the attack to 0.984s after the attack for the 10m/s scenario. When the 2 malicious nodes broadcasts 20 RREQ packets every second, the average end-to-end delay changes from 0.595s before the attack to 4.860 s after the attack for the 30m/s scenario. Figure 9c shows the routing load increases as the flooding attack frequency increases. When the attacker uses 1 malicious nodes and broadcasts 10 RREQ packets every second, the routing load changes from 3.85pkt before the attack to 23.74pkt after the attack for the 10m/s scenario. When the 2 malicious nodes broadcasts 20 RREQ packets every second, the routing load changes from 5.64pkt before the attack to 1,700.94pkt after the attack for the 30m/s scenario.



a) Packet delivery ratio

b) End-to-end delay



c) Routing load

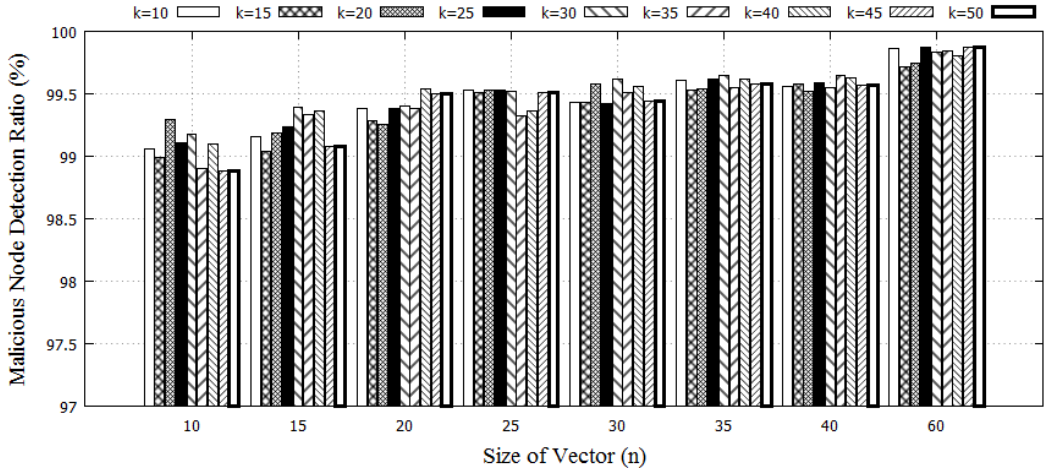
Figure 9. AODV performance under RREQ flooding attacks

4.2.2. Flooding attacks detection performance of FAPRP

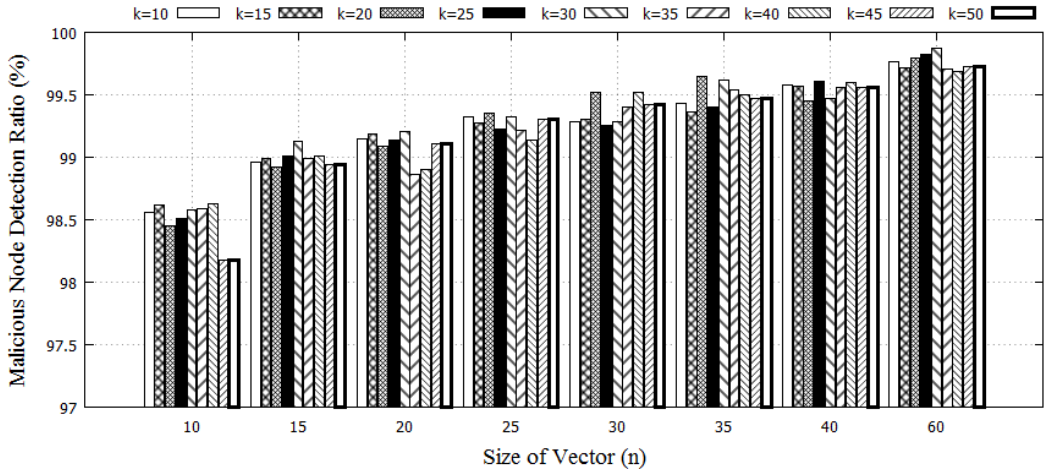
In this section we evaluate the malicious node detection performance of the proposed solution. Malicious node detection ratio is defined in (4). 216 scenarios are

simulated: RDFV of size 10, 15, 20, 25, 30, 35, 40 and 60; the cut off values of k for kNN are set at 10, 15, 20, 25, 30, 35, 40, 45 and 50. Nodes move in a Random Way Point pattern with a specified maximum speed of 10m/s, 20m/s and 30 m/s. 20 source-destination UDP connections are set up among nodes. The intruder uses 2 malicious nodes and floods 20 packets every second.

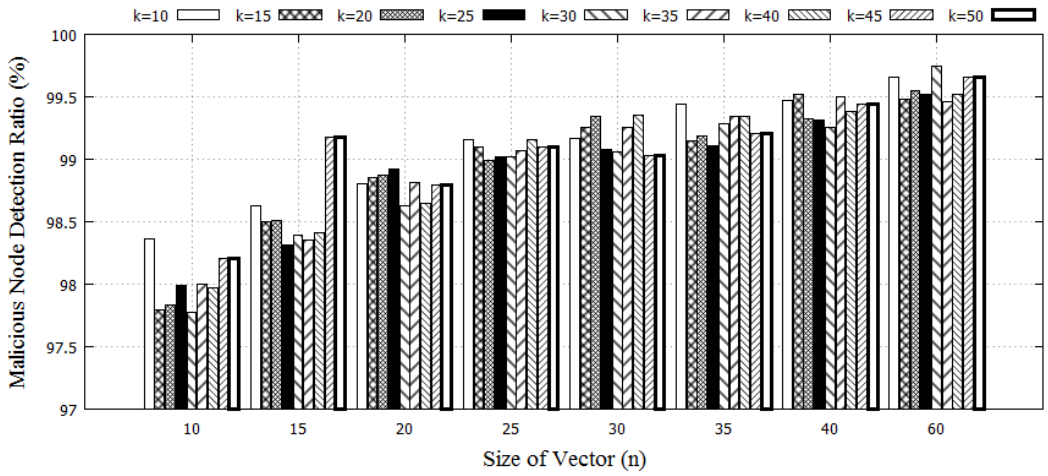
The results in Figure 10 show that by making use of the route discovery history feature vector and the kNN machine data mining algorithm, our method achieves high malicious nodes detection ratio and the complexity of the overall detection algorithm is proportional to the size of the vector. We see that the detection rate of FAPRP is above 99.0% and the mistaken rate is below 1.0% for all scenarios using vector sizes larger than 35. Figure 10d shows that the average of the maximum successful detection rate of FAPRP is above 99.82% when the cutoff value is 30 and vector size is 60. In brief, the proposed solution is effective in detecting the RREQ flooding attacks.



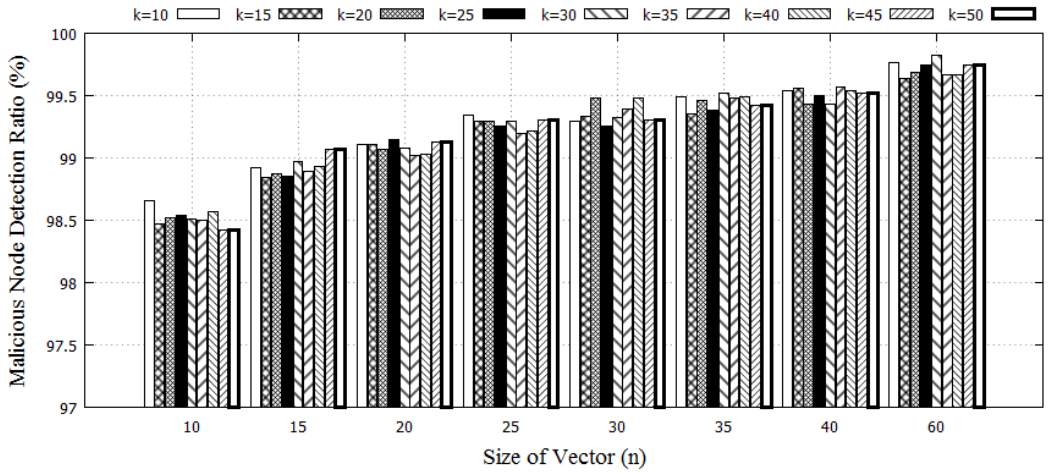
a) 1-10 m/s mobility speed



b) 1-20 m/s mobility speed



c) 1-30 m/s mobility speed



d) Average of mobility speed

Figure 10. Malicious nodes successful detection ratio

4.2.3. Performance evaluation of AODV, B-AODV and FAPRP

In this section we simulate 27 scenarios to evaluate the performance of the AODV, B-AODV and FAPRP protocols under RREQ flooding attacks. The cutoff value (k) is 30 and vector size (n) is 60. All nodes move in a Random Way Point pattern with specified maximum speeds of 10m/s, 20m/s and 30 m/s. 2 malicious nodes, each floods 20 packets every second. 20 pairs of communicating nodes are set up among source nodes. The simulation results are shown in Table 5.

a) Packet Delivery Ratio: The results in Figure 11a show that the average packet delivery ratio for mobility speed by AODV is about 86.79% in the absence of a malicious node. When there is one malicious node, the packet delivery ratio is about 23.17%, and 9.17% for two one malicious nodes. This is due to RREQ flooding of the fake route request packets by the malicious node, resulting in a high consumption of bandwidth and buffer overloads at intermediate nodes with fake RREQs. For B-AODV in normal scenarios, the average packet delivery ratio is about 59.92%. In flooding scenarios, B-AODV average packet delivery ratio is above 56.17% when the

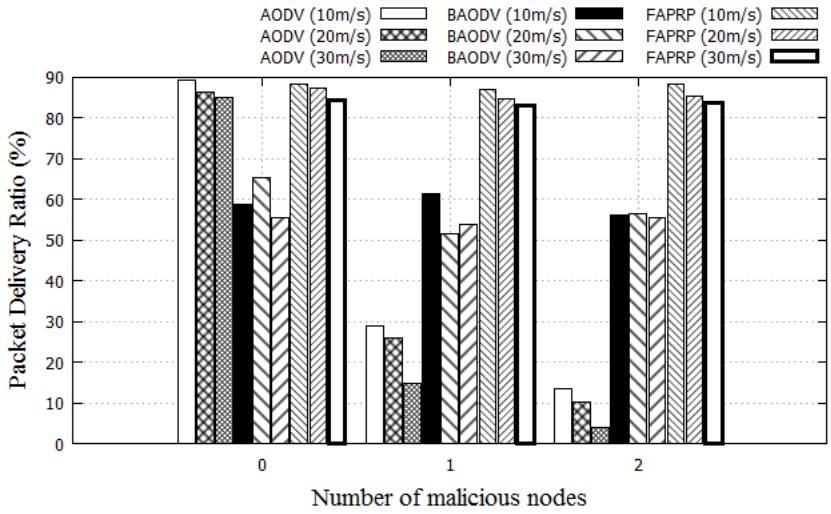
intruder uses one or two malicious nodes. When our proposed solution is deployed, the packet delivery ratio for normal scenarios and high mobility speed is about 86.67%. Under flooding scenarios, FAPRP packet delivery ratio is above 85.02% when the intruder uses one or two malicious nodes. In brief, our solution is more efficient compared to AODV and B-AODV under normal network operation scenarios and more effective in handling RREQ flooding attacks with higher correct detection rates.

Table 5. AODV, B-AODV and FAPRP performances

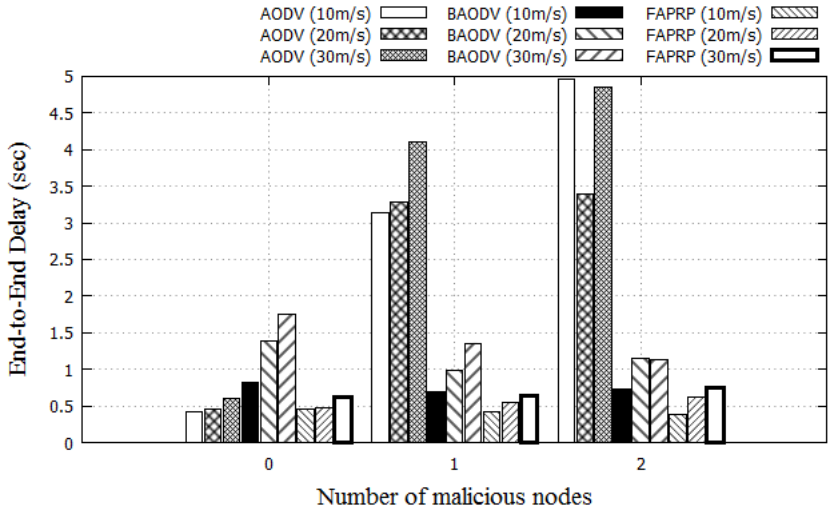
1..10m/s									
	PDR (%)			**RL (pkt)**			**ETE (sec)**		
MN	AODV	BAODV	FAPRP	AODV	BAODV	FAPRP	AODV	BAODV	FAPRP
0	89.19	58.81	88.2	3.85	2.2	3.82	0.420	0.827	0.457
1	28.75	61.39	87.12	139.08	3.32	5.17	3.143	0.689	0.425
2	13.36	56.28	88.25	464.98	5.43	6.15	4.959	0.734	0.386
1..20m/s									
0	86.28	65.39	87.39	4.66	2.98	4.43	0.449	1.384	0.466
1	26.03	51.54	84.75	155.85	3.87	5.88	3.290	0.983	0.548
2	10.33	56.62	85.22	624.38	5.93	6.85	3.397	1.143	0.617
1..30m/s									
0	84.90	55.57	84.43	5.64	2.67	5.44	0.595	1.743	0.622
1	14.74	53.96	83.19	288.98	3.55	6.55	4.108	1.359	0.646
2	3.83	55.62	83.87	1700.94	5.42	7.79	4.860	1.126	0.740
Average									
0	86.79	59.92	86.67	4.72	2.62	4.56	0.488	1.318	0.515
1	23.17	55.63	85.02	194.64	3.58	5.87	3.514	1.010	0.540
2	9.17	56.17	85.78	930.10	5.59	6.93	4.405	1.001	0.581

b) End-to-end delay: The results in Figure 11b show that with AODV, the average end-to-end delay is about 0.488s under normal scenarios. The end-to-end delays are about 3.514s and 4.405s for one and two one malicious nodes respectively. This high end-to-end delay is caused by the broadcasting of selective fake route request packets by the malicious nodes. For B-AODV under normal scenarios, the average end-to-end delay is about 1.138s. Under flooding scenarios, B-AODV end-to-end delay is about 1.010s with one malicious node and 1.001s with two malicious nodes. This is caused by the failure of B-AODV in detecting and preventing flooding attacks resulting in lower packet delivery ratios and longer route discovery delays. For our proposed solution, the average end-to-end delay for normal scenarios and mobility speed is about 0.515s. Under flooding attacks, FAPRP average end-to-end delays are about 0.540s and 0.581s when intruder uses one and two malicious nodes respectively. Clearly, FAPPRP achieves shorter end-to-end delay compared to AODV and B-AODV under both normal and flooding attack scenarios.

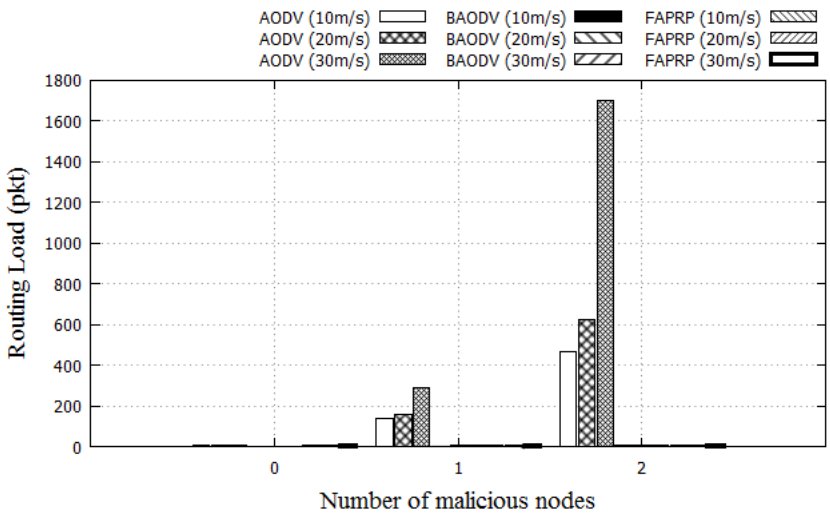
c) Routing load: The results in Figure 11c show that the average routing load for high mobility speed by AODV is about 4.72pkt in the absence of a malicious node. The routing loads are about 194.64pkts and 930.1pkts for one and two one malicious nodes respectively. The high routing load is caused by the broadcasting of selective fake route request packets by the malicious nodes. For B-AODV in normal scenarios, the routing load is about 2.62pkt. B-AODV average routing load in attacks state is about 3.58pkt when the intruder uses one malicious node and 5.59pkt for two malicious nodes. For our proposed solution, the routing load for normal scenario and high mobility speed is about 4.56pkt. Under flooding attacks, FAPRP average routing load is about 5.87pkts and 6.93pkts when the intruder uses one and two malicious nodes respectively. B-AODV routing load is, however, better as compared to AODV as it drops many route request packets due to mistake detection. Overall, FAPRP performs as well as AODV in the routing load measure under both normal and flooding attack scenarios due to its high correct detection rate and low mistake rate.



a) Packet delivery ratio



b) End-to-end delay



c) Routing load

Figure 11. AODV, B-AODV and FAPRP performances under RREQ flooding attacks

5. CONCLUSION

In this paper, we introduced the flooding attack detection algorithm (FADA) based on our proposed route discovery frequency history feature vector and the kNN data mining algorithm to detect and isolate the malicious nodes in the network. We introduced a new FAPRP protocol by integrating FADA into the route request phase of AODV. The simulation results show that FADA achieves malicious nodes successful detection ratio much higher (above 99.0%) than those of existing algorithms, and low mistaken rate (below 1.0%). Furthermore, the proposed solution is efficient in that it improves the network performance in terms of higher packet delivery ratio, smaller end-to-end delay and reduces the routing load compared to AODV and B-AODV protocols.

The limitation of our solution is that it does not deal with spoofing route request flooding attacks and data flooding attacks. In the future, we will extend the proposed solution for mitigating the effects of these attacks.

Data Availability

The underlying data comes from simulation results.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

REFERENCES

[1] H. Jeroen, M. Ingrid, D. Bart, and D. Piet, "An overview of Mobile Ad hoc Networks: Applications and challenges," *Journal of the Communications Network*, vol. 3, pp. 60–66, 2004.

[2] C. E. Perkins, M. Park, and E. M. Royer, "Ad-hoc On-Demand Distance Vector Routing," *In Proceedings of Second IEEE Workshop on Mobile Computing Systems and Applications (WMCSA)*, pp. 90–100, 1999.

[3] C. E. Perkins and P. Bhagwat, "Highly Dynamic Destination Sequence Distance Vector (DSDV) for Moblie Computers," in *Proceedings of the SIGCOMM 1994 Conference on Communications Architectures, Protocols and Applications*, 1994, pp. 234–244.

[4] Z. J. Haas, M. R. Pearlman, and P. Samar, "The Zone Routing Protocol (ZRP) for Ad Hoc Networks," *draftietfmanetzonezrp02 txt*. pp. 1–11, 2002.

[5] E. Alotaibi and B. Mukherjee, "A survey on routing algorithms for Wireless Ad-hoc and Mesh networks," *Computer Networks*, vol. 56, no. 2, pp. 940–965, 2012.

[6] R. Di Pietro, S. Guarino, N. V. Verde, and J. Domingo-Ferrer, "Security in Wireless Ad-hoc Networks - A survey," *Computer Communications*, vol. 51, pp. 1–20, 2014.

[7] R. Mitchell and I.-R. Chen, "A survey of intrusion detection in wireless network applications," *Computer Communications*, vol. 42, 2014.

[8] M. Wazid and A. K. Das, "A Secure Group-Based Blackhole Node Detection Scheme for Hierarchical Wireless Sensor Networks," *Wireless Personal Communications*, vol. 94, no. 3, pp. 1165–1191, 2017.

[9] E. C. H. Ngai, J. Liu, and M. R. Lyu, "An efficient intruder detection algorithm against sinkhole attacks in wireless sensor networks," *Computer Communications*, vol. 30, pp. 2353–2364, 2007.

[10] S. Gurung and S. Chauhan, "A novel approach for mitigating gray hole attack in MANET," *Wireless Networks*, vol. 24, no. 2, pp. 565–579, 2018.

[11] L. Thai-Ngoc and V. Thanh-Tu, "Whirlwind: A new method to attack Routing Protocol in Mobile Ad hoc Network," *International Journal of Network Security*, vol. 19, no. 5, pp. 832–838, 2017.

[12] T. T. Vo, N. T. Luong, and D. Hoang, "MLAMAN: a novel multi-level authentication model and protocol for preventing wormhole attack in mobile ad hoc network," *Wireless Networks*, pp. 1–18, May 2018.

[13] Y. Ping, D. Zhoulin, Y. Zhong, and Z. Shiyong, "Resisting flooding attacks in ad hoc networks," *International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume II*, vol. 2, pp. 657–662, 2005.

[14] H. Ehsan and F. A. Khan, "Malicious AODV: Implementation and analysis of routing attacks in MANETs," *Conference on Ubiquitous Computing and Communications, IUCC-2012*, pp. 1181–1187, 2012.

[15] D. Gada, R. Gogri, P. Rathod, Z. Dedhia, N. Mody, S. Sanyal, and A. Abraham, "A Distributed Security Scheme for Ad Hoc Networks," *ACM Crossroads*, vol. 11, no. 1, pp. 1–14, Sep. 2004.

[16] J. H. Song, F. Hong, and Y. Zhang, "Effective filtering scheme against RREQ flooding attack in mobile ad hoc networks," in *Parallel and Distributed Computing, Applications and Technologies, PDCAT Proceedings*, 2006, pp. 497–502.

[17] P. Yi, Y. Hou, Y. Bong, S. Zhang, and Z. Dui, "Flooding Attacks and defence in Ad hoc networks," *Journal of Systems Engineering and Electronics*, vol. 17, no. 2, pp. 410–416, 2006.

[18] M. J. Faghihniya, S. M. Hosseini, and M. Tahmasebi, "Security upgrade against RREQ flooding attack by using balance index on vehicular ad hoc network," *Wireless Networks*, vol. 23, no. 6, pp. 1863–1874, 2016.

[19] S. Gurung and S. Chauhan, "A novel approach for mitigating route request flooding attack in MANET," *Wireless Networks*, pp. 1–16, 2017.

[20] V. Thanh-Tu and L. Thai-Ngoc, "SMA2AODV: Routing Protocol Reduces the Harm of Flooding Attacks in Mobile Ad Hoc Network," *Journal of Communications*, vol. 12, no. 7, pp. 371–378, 2017.

[21] M. Patel, S. Sharma, and D. Sharan, "Detection and Prevention of Flooding Attack Using SVM," in *2013 International Conference on Communication Systems and Network Technologies*, 2013, pp. 533–537.

[22] W. Li, P. Yi, Y. Wu, L. Pan, and J. Li, "A new intrusion detection system based on KNN classification algorithm in wireless sensor network," *Journal of Electrical and Computer Engineering*, vol. 2014, pp. 1–8, 2014.

[23] DARPA, "The network simulator NS2," 1995. [Online]. Available: http://www.isi.edu/nsnam/ns/.

[24] K. S. Sanjib, K. A. Pankaj, and P. Singh, "Modified K-NN algorithm for classification problems with improved accuracy," *International Journal of Information Technology*, vol. 10, no. 1, pp. 65–70, 2018.

[25] J. Yoon, M. Liu, and B. Noble, "Random waypoint considered harmful," *IEEE INFOCOM 2003*, vol. 2, pp. 1–11, 2003.