Top-K Query Processing And Malicious Node Identification Based on Node Grouping In MANETs

B.Yamini[1], A.Sowmiya[2], M.S.Nandhini[3]
[1]Assistant Professor, Dept of CSE
[2, 3]Dept of CSE
[1, 2, 3] JEPPIAAR SRR Engineering College, Chennai

Abstract- *MANET is Mobile Ad-hoc network in which all the mobile nodes can freely move within the network. In MANET due the query transmissions and the exchange of messages, the traffic will become high. To overcome this top-k query processing is used. This method will retrieve necessary data items among large number of data items. While the process of collecting the information, it may encounter with malicious nodes. Malicious nodes can perform Data Replacement Attack (DRA) or False Notification Attack (FNA). The query issuing node cannot identify the malicious node. So, it inquires all the cluster heads, to narrow down the malicious nodes. Thus, node grouping is achieved by sharing information.*

Keywords- Mobile ad hoc network; Top-k query processing; Node grouping; Data Replacement Attack.

I. INTRODUCTION

A Mobile ad hoc network (MANET) is also known as wireless ad hoc network. It is a continuously self-configuring, infrastructure-less network of mobile devices connected wirelessly. Each node can communicate with other nodes within their network by exchanging data packets. If the source and destination nodes are not in a range, they can transmit packets via intermediate nodes between them. The communication of nodes may take place across the networks.
These MANETS are used in many important areas like rescue operations, military affairs and some other areas. The information retrieval in those large networks will be challenging, need to concern about the mobile node's battery life, bandwidth, etc. An effective method of data retrieval that is top-K query processing is used to retrieve all the necessary data items among the large amount of data items. This is based on the particular attribute scoring. According to this scoring the query issuing node can get the result of necessary nodes with the K-highest scores.

In MANET, a normal node may change into malicious node within the network at any time. The malicious nodes perform some attacks like DRA, FNA. DRA is an attack in which the malicious ode will replace the original data item with unnecessary yet proper data items. FNA will reply with the fake notifications.

To maintain accuracy, nodes send reply messages along multiple routes. To detect DRA, the query issuing node can gain information from the reply messages which contains the specific route that reply messages belong to. By this way the query issuing node can identify the malicious candidate, requesting the proper information for those candidates individually.

II. STATE OF ART

The existing related studies in this section are secure routing top-K query processing methods and reputation systems.

a. SECURE ROUTING METHODS:

The secure routing protocols will include encryption and hashing methods. The request message will be encrypted to send the message securely. This can be done by symmetric key encryption. The reply message from the inquired nodes can be send to the query issuing node by using hashing (Message Authentication Code). Even though implementing these methods, we cannot prevent DRA.

b. TOP-K QUERY PROCESSING:

In proposed methods [4][7], because of the absence of adaptability to node mobility those methods have not been considered. However, those methods can be useful for reducing energy consumption and traffic. In papers [2][3], the authors have been proposed methods for the adaptability of node mobility [5] and it will reduce the traffic and maintain high accuracy.

By retrieving the K-highest scores the query issuing node acquires data items. These scores may be equal to or greater than the threshold value [1] [3]. Although we cannot prevent DRA.

In another method [6], the mobile node sends the data item with the hash value of both priority and superior data item. The source node can check the truthfulness of the received data item by comparing the hash value of the received data item and that of calculated value in the received data item. Even though we cannot handle the DRA.

c. *REPUTATION SYSTEMS*

Reputation systems evaluate the node's performance to remove the malicious nodes which is present in the network. It mostly considers the node's reliability [8], in which each node calculates the local scores of other nodes and pass this information throughout the network [9]. Then the calculation of global reputation score occurs with its own and received local score. Finally, the malicious nodes can be identified by comparing one node's global score with the threshold value where the global score is lower than threshold. In papers, source nodes send their own ID to destination nodes by cryptographic security key. They also attach past and recent reputation scores of destination nodes. Then destination decodes and confirms those scores. Thus, it can prevent the false reputation scores. These methods can share only with source and destination nodes. But we are concentrating with all other nodes.

III. PROPOSED METHOD

a. OVERVIEW:

In this, the query issuing node sends the query to all other nodes in the network and the nodes which receives the information related to the query, stores them on all available and possible routes to that query issuing node. Each receiving node replies to neighbour nodes with K-highest score data items, along with reply message forwarding routes information. These routes contain sender and next node IDs.
The nodes are moving frequently in MANETs, so there is a change in network topology and thus it may lead to radio link disconnection. This disconnection can be rectified by ensuring whether the node in the radio disconnection path can receive queries through another path which was initially sent by the source.

By receiving the reply messages from the nodes, the query issuing node can narrows down the malicious nodes by sending separate inquiries. We can find the malicious nodes which are nearby the query issuing node. But we can't find those nodes which are far away from the query issuing node. That's why the information is shared throughout the network by node grouping based on the similarity of the information between two nodes.

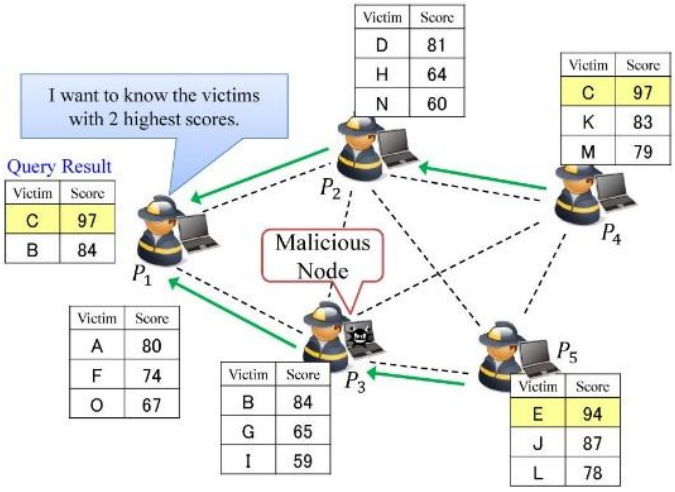

Figure 1: Architecture Diagram

b. TOP-K QUERY PROCESSING:

This includes query forwarding, reply forwarding, link disconnection, detecting attacks.

*1.**Forwarding a query:***

Initially, the query issuing node sends a query over the entire network. The query contains node id of query issuing node, the query identifier of the query, number of requested data items, the condition of the query and list of node identifiers of the nodes along the query path.

The query issuing node denotes the query condition and the number of requested data items. Then the query issuing node transmits a query message whose query path includes its identifier to its neighbour node. The hop count denotes the number of hops to the query issuing node based on the number of nodes included in the query path. Finally, a node which receives the query sets a reply waiting time as follows:

Waiting time for reply= (maximum hop- number of hops) * positive constant.

2.*Reply forwarding:*

When waiting time has paused, each node sends a reply message. It includes its node identifier, next node identifier, a data items list, node id of the data list and a list summarizing the reply message routes.

3.Link disconnection:

When a radio link disconnection to any node (parent or next node) occurs, a replying node cannot send a reply

message. This may result in low accuracy. When a node sends a reply message for a number of times, does not receive an ack from parent or next node, detection of radio link disconnection can be done by the source node. If a source node has no neighbour nodes, it sends a reply message to a neighbour as of reply forwarding process.

4.Detection of attacks:

The query issuing node detects the DRA after it receives the reply messages from all the nodes. We can get the k-highest scores, data list and the route which the reply message is going to be forwarded. If the nodes which have data items in the top-K result are included in the route which contain set of node identifiers from the node having a given data item to query issuing node (Send Route), but the data items in the top-k result are not included in data list (included in the reply message), the query issuing node detects a DRA. It initiates the process of identification of malicious node. If DRA is not detected by the query issuing node, top-k query processing is completed.

c. METHOD OF MALICIOUS NODE IDENTIFICATION:

1. Local Identification:

The query issuing node tries to find the malicious node after exposing a DRA. The query issuing node identifies the malicious node by using Send Route. With the help of the following aspects, we can narrow down the malicious candidates. Set of node id of malicious node candidates (ordered in an ascending with hop count) and replace data items (denoted by missing top-K result). Now the candidates that are listed in the Send Route, considered as possible attackers that is malicious node candidates.

The query issuing node tries to identify the malicious node by sending an inquiry in which the inquiry will present in a route such that the route does not contain any malicious node(MNI-INQ). The nodes which are present near the query issuing node do not receive this inquiry message (hop count=1) because the reply messages from those nodes can be directly received by the query issuing node. That reply messages contain the scores and node id of the data items (candidates in i^{th} position from the query issuing node).

If the reply messages from i-nodes does not contain replaced data items, the query issuing node consider its neighbour node (node position i-1) as malicious candidate and completes the identification process.

2. Notification messages:

After the malicious node identification within the network, this information will be passed each and every node. The process of sending the notification (information) starts from the query issuing node. At first the query issuing node send this notification to its neighbour nodes and those nodes will share this information to other nodes which are present closer to it. Those nodes which already got the notification, ignores it. In this way all nodes share the information.

3. Global identification:

The following two steps are used by each node separately to identify the malicious node (sharing information): node grouping and malicious node identification.

Node grouping:

The network is divided into groups with the help of the information present in the notification messages received by the nodes. Initially the similarity between the normal node and the identified malicious node is calculated. We adopt cosine similarity to reduce the difference in number of identified malicious node. In some groups, there may be a combination of normal nodes and malicious nodes. In order to prevent this, a node performs a cleaning method and it depreciates inconsistency. When a node identifies another node as malicious node it eliminates from the group as well as when a node identifies another node which is less identified when comparing to other nodes in the same group which also eliminated.

By this way we can prevent the invasion of malicious node and fake information. This achieves high similarity.

Malicious node identification:

The malicious nodes can be determined based on information about them which are identified by nodes in each group. There are three kinds of groups: i) consists of only normal nodes, ii) only malicious nodes, iii) combination of both. We can easily identify the malicious node in (i) and (ii) but it is difficult to identify them in a group of (iii), which perform FNA.

We cannot surely tell about the presence of malicious nodes. We consider they are present in minorities thus the suspected nodes are conformed as malicious when they are detected as malicious by a certain number of other groups which is greater or equal to certain threshold.

Table.1: Scores of the Data Items

NODE	SOURCE							
M1	79	72	69	56	55	47	32	29
M2	72	65	62	59	51	49	40	22
M3	95	76	75	61	53	46	37	35
M4	84	81	79	71	66	60	58	27
M5	91	80	77	54	44	36	25	19
M6	98	86	78	67	58	42	38	30

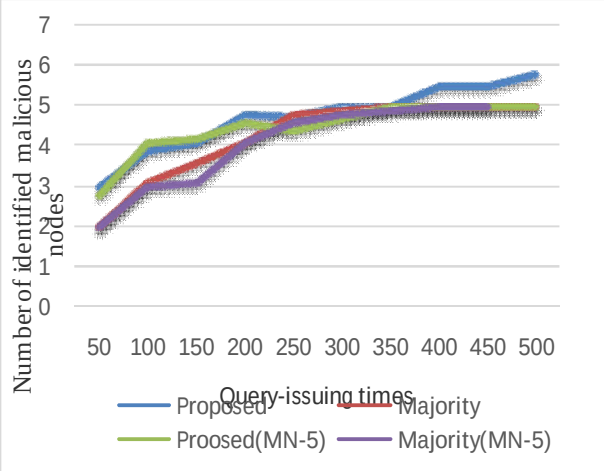

Figure 2: Effect of Number of queries. -Number of identified malicious node.

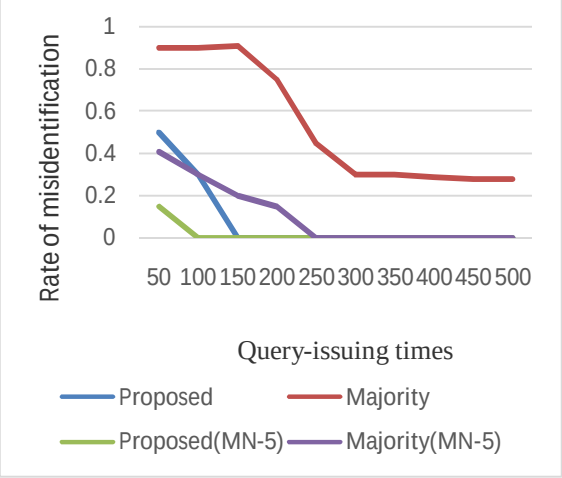

Figure 3: Effect of Number of queries – Rate of the misidentification.

IV. CONCLUSION

The proposed method is top-k query processing and malicious node identification based on node grouping in MANETs. In order to maintain high accuracy of the query result and detect attacks, nodes reply with k data items with the highest score along multiple routes. After detecting attacks, the query-issuing node narrows down the malicious node candidates and then tries to identify the malicious nodes through message exchanges with other nodes. When, multiple malicious nodes are present, the query issuing node may not be able to identify all malicious nodes at a single query. It is effective for node to share the information about the identified malicious nodes with other nodes. In this method, each node divides all nodes into some groups by using the similarity of the information about the identified malicious nodes. Then, it identifies malicious nodes based on the information on the groups.

REFERENCES

[1] T. Tsuda, Y. Komai, Y. Sasaki, T. Hara, and S. Nishio, ‘‘Top-k query processing and malicious node identification against data replacement attack in MANETs,’’ in Proc. MDM, Jul. 2014, pp. 279–288.

[2] D. Amagata, Y. Sasaki, T. Hara, and S. Nishio, ‘‘A robust routing method for top-k queries in mobile ad hoc networks,’’ in Proc. MDM, Jun. 2013, pp. 251–256.

[3] Y. Sasaki, T. Hara, and S. Nishio, ‘‘Two-phase top-k query processing in mobile ad hoc networks,’’ in Proc. NBiS, Sep. 2011, pp. 42

[4] B. Chen, W. Liang, R. Zhou, and J. X. Yu, ‘‘Energy-efficient top-k query processing in wireless sensor networks,’’ in Proc. CIKM, 2010, pp.329–338

[5] R.Hagihara,M.Shinohara,T.Hara,andS.Nishio,‘‘A message processing method for top-k query for traffic reduction in ad hoc networks,’’ in Proc. MDM, May 2009, pp. 11–20

[6] J. Shi, R. Zhang, and Y. Zhang, ‘‘Secure range queries in tiered sensor networks,’’ in Proc. INFOCOM, Apr. 2009, pp. 945–953

[7] W.-T. Balke, W. Nejdl, W. Siberski, and U. Thaden, ‘‘Progressive distributed top-k retrieval in peer-to-peer networks,’’ in Proc. ICDE, Apr. 2005, pp. 174–185.

[8] M.Srivatsa,L.Xiong,andL.Liu,‘‘TrustGuard :Countering vulnerabilities in reputation management for decentralized overlay networks,’’ in Proc. WWW, 2005, pp. 422–431.

[9] S. Buchegger and J.-Y. Le Boudec, ‘‘Performance analysis of the CONFIDANT protocol,’’ in Proc. MobiHoc, 2002, pp. 226–236

[10] T. Camp, J. Boleng, and V. Davies, ‘‘A survey of mobility models for ad hoc network research,’’ Wireless Commun. Mobile Comput., vol. 2, no. 5, Sep. 2002, pp. 483–502.
[11] A. D. Wood and J. A. Stankovic, ‘‘Denial of service in sensor networks,’’ Computer, vol. 35, no. 10, Oct. 2002, pp. 54–62
[12] S. J. Lee and M. Gerla, ‘‘Split multipath routing with maximally disjoint pathsinadhocnetworks,’’inProc.ICC,vol.10.Jun.2001,pp.3 201–3205