ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 7, Issue 5, May 2018

Top-k query processing and malicious node identification in MANETS using node grouping

Amaya Rose Abraham, R.Kannan

***Abstract*— In mobile ad hoc networks (MANETs), it is effective to retrieve data items using top-k query. We propose methods for top-k query processing and malicious node identification based on node grouping in MANETs. In order to maintain the accuracy of the query result, nodes reply with k data items with the highest score along multiple routes, and the query-issuing node tries to detect attacks from the information attached to the reply messages. After detecting attacks, the query-issuing node 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 a node to share information about the identified malicious nodes with other nodes. In our method, each node divides all nodes into 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. The simulation experiments are done by using a network simulator with high accuracy of the query result.**

Keywords— Adhoc networks, MANETs, OLSR, FNA, Data replacement attack

I. INTRODUCTION

Recently, there has been an increasing interest in mobile ad hoc network (MANET), which is constructed by only mobile nodes. Since such self-distributed networks do not require pre-existing base stations. In MANETs, since each node has poor resources it is effective to retrieve only the necessary data items using top-k query, in which data items are ordered according to a particular attribute score, and the query-issuing node acquires the data items with k highest scores in the network. Malicious nodes attempt to disrupt query-issuing node's acquisition of the global top-k result for a long period, without being detected. a remarkable characteristic of top-k query processing is that the query-issuing node does not know the global top-k result beforehand. Therefore, even if a malicious node replaces high-score data items with its own low-score ones, when relaying the data items, it is difficult for the query-issuing to detect the attack, and it may believe that all the received data items with k highest scores are the global top-k result.

Manuscript received May, 2018.

***Amaya Rose Abraham**, Electronics and communications Department, Anna University Chennai / R.V.S College of engineering and technology Coimbatore/ Cimbatore, India, 9496665868*

II. EXISTING SYSTEM

A. *Mobile ad hoc networks*

MANET is a collection of mobile devices connected through wireless links to serve a specific purpose. MANETs provide users with easier ways to connect and communicate without the need for prior setup or a centralized server. MANETs are particularly used in situation where a fast installation is needed and no infrastructure is available. The only condition the device has to fulfill is the communication interface, as it needs one to build up a connection to other devices. The networks are self-organized and adaptive. As it has no infrastructure the participants are directly connected with one another and not to an access point, to a gateway or something similar.

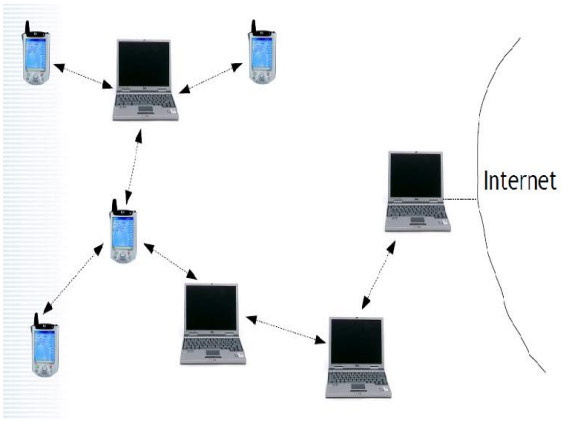

Fig. 1 MANET

B. *Attacks in MANET*

Many characteristics might be used to classify attacks in the ad hoc networks Passive attacks are launched to steal valuable information in the targeted networks. Examples of passive attacks in ad hoc network are eavesdropping attacks and traffic analysis attacks. Passive attacks do not intend to disrupt the network operations, active attacks on the other hand actively alter the data with the intention to obstruct the operation of the targeted networks. Examples of active attacks comprise actions such as message modifications, message replays, message fabrications and the denial of service attacks. The attacks in the MANETs due to the malicious nodes are basically classified into two types. They are: (1). Data Replacement Attack (DRA) and (2). False

Notification Attack (FNA). DRA is a new type of attack in which a malicious node replaces the received data items known as the local top-k result with unnecessary data items for example its own low-score data items. Since DRAs are a strong attack, and more difficult to detect than other traditional types of attacks, some specific mechanism for defending against DRAs are required. . A malicious node may declare fake information that claims normal nodes as the malicious nodes is called as the FNA. The one who is doing only the attack called FNA is known as the liar nodes where it is represented as LN.

The main objective of the paper is that addressing the issue of identification of liar nodes (LNs), and designing a message authentication method using the OLSR to prevent malicious nodes from performing FNA in MANETs.

C. *Disadvantages of existing system*

• The current mobile ad-hoc networks allow for many different types of attacks.

• Current MANETs are basically vulnerable to two different types of attacks: active attacks and passive attacks.

• Active attack is attack when misbehaving node has to bear some energy costs in order to perform the threat.

• Passive attacks are mainly due to lack of cooperation with the purpose of saving energy selfishly.

• Nodes that perform active attacks with the aim of damaging other nodes by causing network outage are considered as malicious while nodes that make passive attacks with the aim of saving battery life for their own communications are considered to be selfish.

III. PROPOSED SYSTEM

The data replacement attack (DRA), in which a malicious node replaces the received data items (which we call the local top-k result) with unnecessary yet proper data items (e.g., its own low-score data items). Since DRAs are a strong attack, and more difficult to detect than other traditional types of attack, some specific mechanism for defending against DRAs are required. We propose top-k query processing and malicious node identification methods again DRAs in MANETs. In the top-k query processing method, in order to maintain accuracy of query result and detect attacks, nodes reply with data items with k highest scores along multiple routes. Moreover, to enable detection of DRA, reply messages include information on the route along which reply messages are forwarded, and thus the query-issuing node can know the data items that properly belong to the message. In the malicious node identification method, the query-issuing node first narrows down the malicious node candidates, using information in the received message, and then requests information on the data items sent by these candidates. In this way, the query-issuing node can identify the malicious node. When there are multiple malicious nodes in the network, it is difficult to identify all the malicious nodes in a single query. By using our methods, nodes are likely to identify the malicious nodes which are near their own location, while they hardly identify the malicious nodes which are far from their own location. Therefore, in order to quickly identify more malicious nodes, it is effective to share the information about the identified malicious nodes with other nodes. In this case, however, a malicious node may declare fake information that claims normal nodes as the malicious nodes (false notification attack (FNA)).We need some method to correctly identify the malicious nodes against FNAs.

Therefore, in our malicious node identification method, after nodes share the malicious node identification information, each node divides all nodes into some groups based on the similarity of the information. Then, the node determines the final judgment of malicious nodes based on the judgment result of each group. In our method, even if malicious nodes claim that normal nodes are the malicious nodes, there is a decisive difference in the nature of the information possessed by normal and malicious nodes concerning the identified malicious nodes, and therefore, the normal nodes can easily identify the malicious nodes. Furthermore, even if malicious nodes mix the correct information on malicious nodes identified by other normal nodes with their fake information, in order to increase their similarity with normal nodes, the normal nodes in the same group will nonetheless certainly identify the malicious nodes, but not normal nodes. Thus, the information from the malicious nodes can be removed and there is little influence of FNAs.

A. *Top-k query processing methods*

In the field of database systems and distributed systems, top-k query is effective to retrieve only the required data items in a large amount of data items. However, the other methods do not protect against DRA, and are unsuitable for use in MANETs, because they are not adapted to node mobility. We proposed top-k query processing methods for MANETs, adapted to the node mobility, maintaining high accuracy of the top-k result and reducing traffic. The query messages include the scores of data items, and nodes narrow down candidates that may include the global top-k result, resulting in reduced communication traffic volume. The query- issuing node first retrieves the k-highest score (threshold) in the network, and then acquires data items with scores equal to or greater than the threshold. each sensor node sends each data item attached both the hash value of one priority data item and that of one superior data item. After the source node received the top-k result, it ensures the safety of the received data items to check whether the received hash values correspond with hash values calculated by the received data items. In these methods, the sender node protects against fabrication of data items by sending data items encrypting with a symmetric key. Malicious nodes

generate new and false data items (i.e., other nodes' data items or data items whose score are not same as the score calculated from raw data items and query conditions) and send back them. However, we assume that raw data items are generated from some special devices and software such as medical sensors, which can be read but cannot be modified even by the owner nodes. Therefore, we assume that malicious nodes perform DRAs, which malicious nodes replace necessary data items with unnecessary yet proper data items.

B. Assumptions

The system environment is assumed to be a MANET constructed by mobile nodes held by members of a highly important collaborative work such as rescue operations and military affairs in which the members issue top-k queries to efficiently acquire data items. In a case of rescue operations, ambulance clews need to pick up victims in a critical condition. The attackers such as terrors hack a node which an ambulance clew holds because the attackers aim to spread the damage for a long time. The ambulance clew whose node has been hacked does not recognize that his/her own node has been hacked, and the malicious node sends data items which he/she does not intend.

C. System model

The set of all mobile nodes in the system is denoted by M .D = {M1,M2; : : : ;Md}, where n is the total number of mobile nodes and Mi($1 \leq i \leq n$) is a node identifier. The set of all data items in the entire network is denoted by D. D = {D1,D2; : : : ; Dd}, where d is the total number of data items and Di($1 \leq i \leq d$) is a data identifier. Each data item is retained by a specific node. Since we assume a highly important collaborative work, highly secure communications and data exchanges are essential. Therefore, we assume that each node has the public key of all nodes in the network. When a node replies with data items (i.e., sends a reply message), it encrypts the data items using the public key of the destination node to avoid intermediate nodes modifying and reading the data items. In addition, the node sends the reply message (including the encrypted data items) to some of its neighbors after encrypting the message using the neighbors' public key. This is to ensure a secure communication with neighbors and avoid others overhearing the message. On the other hand, when a node sends (or relays) a query, it broadcasts the query without encryption since a query is not aimed to send to specific nodes but should be sent to all neighbors. The scores of data items can be calculated based on the query condition and specified scoring functions. Raw data items are generated from some special devices and software (which are independent of the mobile nodes' OS and applications) such as medical sensors, which can be read but cannot be modified even by the owner nodes. Therefore, even if a node is hacked by attackers, it cannot modify its own data items, i.e., a malicious node cannot generate incorrect and fake data items whose scores are unfairly high. This assumption is also to achieve highly secure collaborative work. In order to acquire data items with the k highest scores in a MANET, each intermediate node should selectively send data items with higher scores. Therefore, the scores of reply data items in a reply message are not encrypted, with the public key of the query-issuing node, i.e., each mobile node can know the scores of data items in the reply message.

D. Attack model

We assume that the number of the malicious nodes in the network is m. A malicious node seeks a way to disrupt the query-issuing node's acquisition of the global top-k result, without being detected. If the malicious node falsifies the scores of its own data items or that of others own data items when relaying them, the query-issuing node can easily detect the attack by comparing received data items' scores (attached directly with the reply message) with the scores calculated from the received data items. Thus we assume that malicious nodes attempt only a DRA in top-k query processing. When a malicious node does DRA, it randomly replaces d [h . k] (i.e., h denotes the rate of replacement) data items in the local top-k result with its own data items, which have lower scores than the local top-k result. Moreover, we assume that each node floods the entire network with the information about the identified malicious nodes in order to share it with other nodes. Aiming to confuse normal nodes and make them misjudge the malicious node identification, a malicious node does a FNA, where it notifies some normal nodes as malicious nodes. If each malicious node randomly notifies normal nodes as malicious nodes, the FNA is easily detected by other normal nodes because only this node claims these normal nodes as malicious ones in most cases. Therefore, multiple malicious nodes collaborate to notify same normal nodes. This type of FNA has more influence in malicious node identification. We assume that malicious nodes do two types of attacks (i.e., DRA and FNA), but do not always do them, i.e., they sometimes do only one type of attack and sometimes do both.

IV. MALCIOUS NODE IDENTIFICATION METHODS

A. Local identification

After detecting a DRA, the query-issuing node tries to identify the malicious nodes. The query-issuing node narrows down the candidates for malicious nodes, and identifies the malicious nodes by making respective inquiries. In our proposed method the query-issuing node narrows down the malicious node candidates by using SendRoute Candidate denotes the set of node identifiers of malicious node candidates, ordered by ascending hop count from the query-issuing node, and missing Top-k result denotes the replaced data items. The nodes included in SendRoute, whose data items are corrupted (by the malicious node), are all possible attackers. Therefore, the query-issuing node recognize these nodes as malicious node candidates.

When the number of malicious node candidates is one, the query-issuing node identifies this node as the malicious node and completes the procedure.

B. Sending notification message

After identifying the malicious nodes, the query-issuing node floods the information on the identified malicious nodes within the network. More specifically, the query-issuing node, Mp, sends a notification message to its neighboring nodes. The notification message contains the query identifier of the query (QNum), the node identifier of the query-issuing node (Mp), and the list of the node identifiers of the identified malicious nodes (BLp). The node, which received the notification message, stores the message, and also forwards it to the neighboring nodes. The node, which received the same notification message again, ignores the message, and also forwards it to the neighboring nodes. Hence, all nodes share the information on the identified malicious nodes in the network.

C. Global identification

In this method, each node individually identifies malicious nodes using the shared information by the two steps; node grouping and malicious node identification.Node Grouping: Each node divides nodes in the network into some groups based on the information in the notification messages received by the nodes each node starts this process (i.e., grouping) after receiving Numquery queries. R(i =1; 2; : : : ; n) denotes the evaluation score by Mi, which is represented by an n-dimensional vector, and indicates the malicious nodes identified by Mi. More specifically, the j-th element of Ri (j=1; 2; : : : ; n) is set to 1 when Mi identified Mj as the malicious node, and 0 otherwise. sim(a; b) denotes the similarity of evaluation scores between Ma and Mb. Group denotes groups determined by node grouping, Gcan denotes candidates of groups, and Groupg denotes the g-th group in Group. BLg denotes malicious nodes identified by nodes in Groupg, Mg,e denotes a node in Groupg and CountBLg,f denotes the number of nodes which identify Mf included in BLg as a malicious node among nodes in Groupg. θ denotes the threshold for the grouping, and ρ denotes the threshold for the cleaning, which is represented by $\rho = |Group_g| \cdot \alpha$ (Here, $|Group_g|$ denotes the number of nodes in Groupg, and α denotes a system parameter ($0 \leq \alpha \leq 1$)). First, each node calculates the similarity of nodes in terms of identified malicious nodes based on the received notification messages. In order to decrease the influence of differences in the number of identified malicious nodes among nodes. After the node grouping, some groups may include both normal and malicious nodes. Therefore, the node performs a cleaning in each group to remove the inconsistency. Specifically, if a node, Mg,e, in a certain Groupg, identifies another node in the same group as a malicious node, Mg,e is eliminated from Groupg After that, a node identifying another node which is identified by less than a certain number of nodes in the same group, is also eliminated from the group

D. Cases of not detecting DRA

In our top-k query processing method, each node sends back data items to two neighbor nodes, and the query-issuing node successfully acquires data items in the top-k result, even if one of the routes includes a malicious node, because the alternate route can safely ensure that the required data items are properly sent back. However, especially when the node density in the network is low, some nodes may not have multiple neighbor nodes, and can send back data items along only one route. If data items are sent through a malicious node on the singular route or all two nodes on multiple paths are malicious, the query-issuing node will not acquire data items replaced by the malicious node. Moreover, in our proposed method, the query-issuing node can detect attacks only when it receives reply messages from multiple nodes. For example, when the query-issuing node has only one neighbor node, it cannot detect attacks.

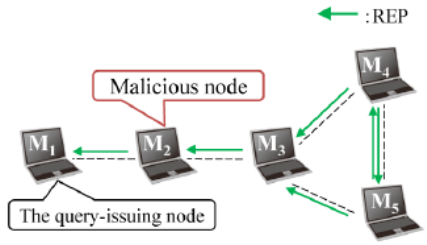

Fig. 2 Unrecognized attack.

In figure 2 the query-issuing node, M1, receives a reply message only from M2, and thus cannot recognize the DRA.

In figure 3 for example, since the malicious node, M4, does not receive reply messages from any other nodes, it cannot attack, and the query-issuing node can acquire the correct top-k result. In this case, the malicious node sends normal reply messages, because other nodes may recognize the node as malicious if it ignores the message.

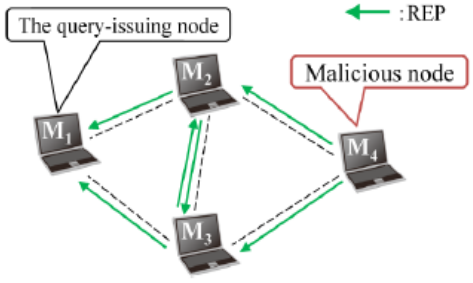

Fig. 3 Disabled attack

In figure 4 though the malicious node, M4, replaces data

items, the corrupted local top-k result is not included in the global top-k result. Thus, the global top-k result is not affected by the DRA. Of course, in these cases, though the query- issuing node can acquire the data items in the global top-k result, it cannot detect the DRA or identify the malicious node.

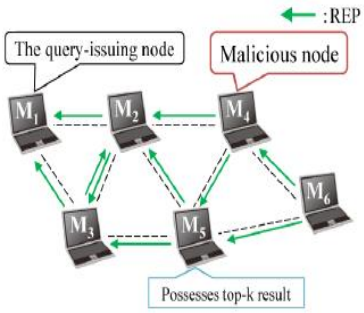

Fig. 4 Ineffective attack

E. Influence of FNA

In the global identification method, each node identifies malicious nodes based on the shared information on the identified malicious nodes. However, the malicious nodes attempt FNAs to configure normal nodes and make them misjudge the malicious node identification. In addition, to disturb the identification, some malicious nodes may do only FNAs (we call them liar nodes). In this section, we discuss the influence of FNA.

First, when a malicious node notifies the information on a randomly selected node as a malicious node, there is little influence of the identification. This is because only few malicious nodes claim the same normal node as malicious while other (many) nodes do not. Therefore, even by a majority based method, where each node identifies a node as malicious when the number of nodes identifying it is more than a threshold, the malicious nodes are substantially identified. Our proposed method can also defend against FNAs, since the similarity among normal nodes and malicious (or liar) nodes is low, i.e., malicious nodes have little possibility to be classified into the same groups with normal nodes, and the number of groups consisting of normal nodes is generally much more than that of malicious nodes. Therefore, there is little possibility to determine normal nodes as malicious nodes.

Next, some malicious and liar nodes may collaboratively claim the same normal node as a malicious node. In this case, because the number of nodes which notify a normal node as a malicious node becomes large, by a single majority-based method, the normal node may be conclusively determined as a malicious node. Here, it should be noted that normal nodes tend to identify near-by malicious nodes, and thus, the identified malicious nodes have some diversity among them. In our proposed method, since nodes which identify the same nodes as malicious are usually classified into the same groups, the number of groups including malicious nodes which have done FNAs is small. Therefore, the misidentification caused by FNAs is less happened than that in the simple majority method. Moreover, malicious nodes may try to increase their similarity with other nodes (for example, by announcing information that includes nodes identified by normal nodes as malicious), in order to increase the number of groups that include them. In our proposed method, only nodes identified by all nodes in each group are decided as malicious nodes identified by the group. Therefore, there is little possibility to conclusively determine normal nodes as malicious nodes.

V. SIMULATIONS

The most creative and challenging phase of the life cycle is system design. The term design describes a final system and the process by which it is developed. The designer's goal is how the output is to be produced and in what format. Samples of the output and input are also presented. Second input data and database files have to be designed to meet the requirements of the proposed output. The general objective is to make information necessary, quick, inexpensive and flexible for the user. Database allows the data to be protected and organized separately from other resources. A database is a collection of interested data stored with minimum redundancy to serve many users quickly and efficiently. The general objectives considered in database design are controlled redundancy, case of learning and use, data independency, more information at low cost, accuracy and integrity, recover from failure, privacy and security performance.

The work is done in Microsoft.NET software and Microsoft .NET is a revolutionary multi-language platform that knits various aspects of application development together with the internet. All .NET programs are compiled to an Intermediate Language (IL) rather than to native code, which can be understood by the computer processor. MSIL must be converted to CPU-specific code by a Just in time (JIT) compiler. It helps the runtime and garbage collection keep track of memory that will be released back to the operating system when it is no longer needed. Microsoft's .net is the next generation platform for building web applications and web services.

The Fig. 5 shows the prevention of MANETs from malicious node performing FNA. It contains 50 nodes and the node 5 is the query issuing node. The transmission range is between100 to 900. If we start the process it locally identifies and node grouping is done.

The Fig.6 shows the node grouping and local identification. The FNA is detected in this step. The grouped nodes are shown in the box.

The Fig.7 shows the narrowing and local identification of data and Fig. 8 detects the nodes. The detected nodes is identified and shown in dialogue box.

The top k results are founded and the data items and reply results are shown in Fig. 9.The top k reply, data items are shown in different dialogue boxes.

ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 7, Issue 5, May 2018

Fig. 5 Nodes formed

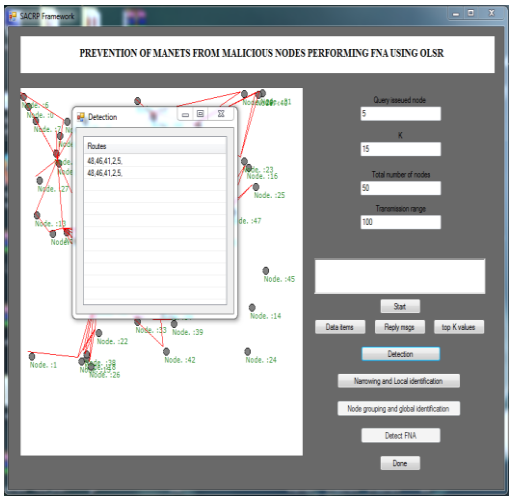

Fig. 8 Detection

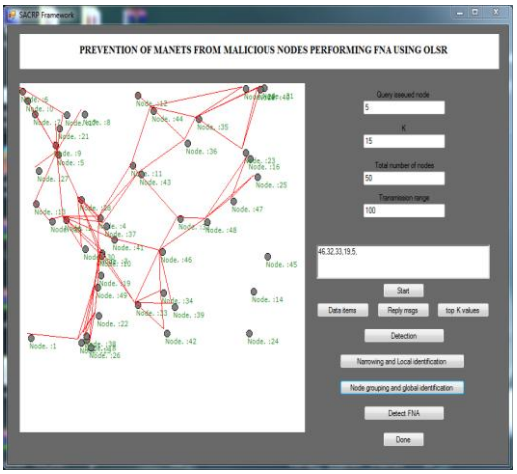

Fig. 6 Node grouping and local identification

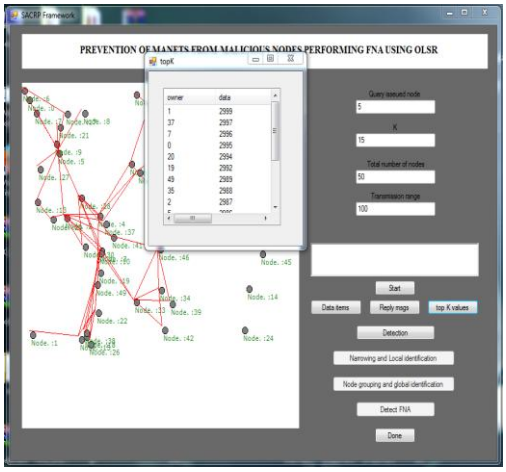

Fig. 9 Top k data

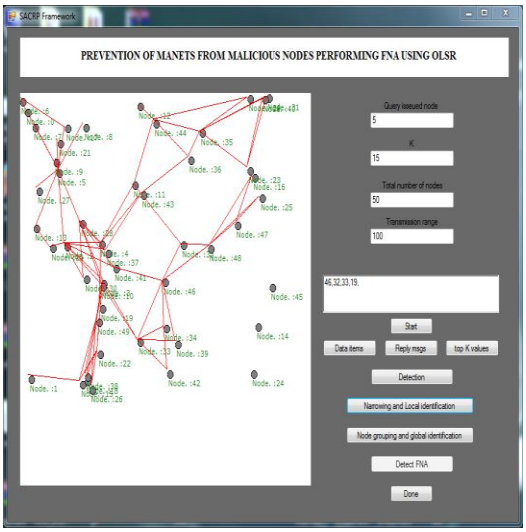

Fig. 7 Narrowing and local identification

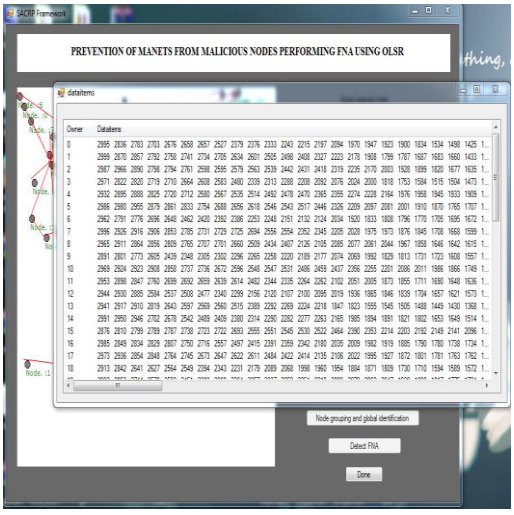

Fig. 10 Reply results

VI. CONCLUSIONS

The proposed methods for 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 our 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] D. Amagata, Y. Sasaki, T. Hara, and S. Nishio, (Jun. 2013) 'A robust routing method for top-k queries in mobile ad hoc networks' in Proc. MDM, , pp. 251_256.

[2] W.-T. Balke, W. Nejdl, W. Siberski, and U. Thaden, (Apr. 2005) 'Progressive distributed top-k retrieval in peer-to-peer networks,' in Proc. ICDE, , pp. 174_185.

[3] S. Buchegger and J.-Y. Le Boudec, (2002) ' Performance analysis of the CONFIDANT 99protocol,' in Proc. MobiHoc, , pp. 226_236.

[4] T. Camp, J. Boleng, and V. Davies, (Sep. 2002) 'A survey of mobility models for ad hoc network research,' Wireless Commun. Mobile Computing., vol. 2, no. 5, pp. 483_502.

[5] B. Chen, W. Liang, R. Zhou, and J. X. Yu, (2010) ' Energy-ef_cient top-k query processing in wireless sensor networks,' in Proc. CIKM, , pp. 329_338.

[6] H. Chan, A. Perrig, and D. Song, (2006) 'Secure hierarchical in-network aggregation in sensor networks,' in Proc. CCS, pp. 278_287.

[7] S. Chen, Y. Zhang, Q. Liu, and J. Feng, (Nov. 2012) 'Dealing with dishonest recommendation: The trials in reputation management court,' Ad Hoc Netw., vol. 10, no. 8, pp. 1603_1618.

[8] Takuji Tsuda, Yuka Komai, Takahiro Hara, Shojiro Nishio (2016), 'Top-k Query Processing and Malicious Node Identification Based on Node Grouping in MANETs 'IEEE Trans. Mobile Computing, no. 7, pp. 2541864.

[9] B. Malhotra, M. A. Nascimento, and I. Nikolaidis(2011), "Exact top-k queries in wireless sensor networks," IEEE Trans. Knowl. Data Eng., vol. 23, no. 10, pp. 1513_1525.

[10] M. Wu, J. Xu, X. Tang, and W. C. Lee (2007), 'Top-k monitoring in wireless sensor networks,' IEEE Trans. Knowl. Data Eng., vol. 19, no. 7, pp. 962_976.

[11] Hagihara, M. Shinohara, T. Hara, and S. Nishio (2009), 'A message processing method for a top-k query for traffic reduction in ad hoc networks,'Proc. Int. Conf. on Mobile Data Management, pp. 11–20.

[12] Ashish Kumar, Vidya Kadam, Subodh Kumar and Shital Pawar(2011), 'An Acknowledgement- Based Approach for the Detection of Routing Misbehavior in MANETS ' in Proc IJAES Volume 1, Issue 1, pp-04-06.

[13] W. Lou, W. Liu, and Y. Fang, (2004) 'SPREAD: Enhancing data con_dentiality in mobile ad hoc Networks,' in Proc. INFOCOM, vol. 4. Mar,pp. 2404_2413

[14] M. Kefayati, H.R. Rabiee, S.G. Miremadi, and A. Khonsari , (2006) 'Misbehavio Resilient Multi-Path Data Transmission in Mobile Ad-Hoc Networks', Proc. Fourth ACM Workshop Security of Ad Hoc and Sensor Networks (SASN '06)

[15] P. Andreou, D. Zeinalipour-Yazti, M. Andreou, P. K. Chrysanthis, and G. Samaras. Kspot (2009) ' Effectively monitoring the k most important events in a wireless sensor network'. In Proceedings of IEEE ICDE (demo)

Amaya Rose Abraham doing M.E in Communication systems in (Electronics and communications department) R.V.S College of engineering and technology Coimbatore. Completed B.tech in Electronics and communications in Ahalia school of engineering Palakkad.

R.Kannan pursing Ph.D and working as assistant professor in R.V.S college of engineering and technology Coimbathore.