International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 25 Issue 4 – APRIL 2018.

An Efficient Top-K Query Processing in MANETS on Malicious Node Identification

A.Mahendra[#1], K..R. Harinath[*2]and M.Shalima Sulthana[#3]

[#]*Dept of CSE, Guest Faculty In IIIT RGUKT RKVALLEY, Vempalli, India*

[*] *Dept of CSE, Assistant Professor In RGMCET, Nandyal, India*

***Abstract*— In mobile ad hoc networks (MANETs), data items can be retrieved effectively by using topk query processing method. The environment which contains malicious nodes cannot provide an accurate result. This paper assumes that the malicious node tries to attempt an attack called data replacement attack in which the necessary data sets are replaced by unnecessary data sets. The proposed work includes node grouping method in top-k query processing for detecting the malicious node. The accuracy of the query result can be maintained by forwarding the data sets along multiple routes and based on the information attached to the reply message the query-issuing node can detect the attack. By exchanging the message, single malicious node can be identified. To identify multiple malicious nodes, it's necessary to share the information of identified malicious node to other nodes. In this method, nodes are grouped based on the similarity of the identified malicious node. Simulation experiments are conducted by using a network simulator, NS2, to verify that this method provides high accuracy and identifies multiple malicious nodes**

***Index Terms*— Mobile ad hoc networks, query processing, routing, traffic, data replacement attack, node grouping.**

I. INTRODUCTION

A self-configuring and infrastructure- less mobile node with wireless link is called mobile ad hoc network. Each node can move independently and capable of routing traffic to other nodes in the network and these network don't have fixed infrastructure. To route the packet each node have to maintain the routing information and it has to be updated when the topology changes. In a MANET, nodes with in communication range can communicate directly with other node, if a node is outside the communication range it has to relay on some other nodes to forward the date. The implementation of MANETs is more relabel because no infrastructure is needed, the network can be deployed in emergency situation. Due to technology improvement MANET enters the area like distribute computing, gaming etc. The mobile ad hoc network has gained momentum recently due the nature of self-distribution of nodes without a base station. In MANETs, each node has limited resource like communication bandwidth, batter life so they depend on other nodes for effective query processing. The nodes will use a top-k query processing method to get the knowledge of the entire network. A node will issue a query this query will moves form one node to another node based on the routing algorithm, and each node will contribute their answer and the items are ordered according to their attribute score and the query-issuing node will get the date with the k highest score in the network. In this environment some node will be a malicious tries to confuse and disrupt the normal operation of the network. The malicious activity can be of two fold the first case the malicious nodes will deploy Data Replacement Attack (DRA), the second one is False Notification Attack (FNA) .The network with malicious nodes will continue to operate normally the user of the network will continue to operate without the knowledge of the presence of the malicious node. The malicious node will try to implement Denial of Service attack and blocks the query processing inside the network. In the case of top-k query processing the some scoring function is used to identify the top-k object the object score depend on the object characteristic like price, color of a product, and cost of the product in online shopping database. The Data objects manipulated based on the scoring predicates that contributes to the total score value. Due to increase in research activity in the area of top-k query processing, the impact of efficient top-k processing is becoming evident in an increasing number of applications [1]. The malicious node tries to disrupt the query issuing node by denying the global top-k query result for long period with being noticed this is called as denial of service attack. The DoS attack is being studied by many researcher for long time so it has many techniques as counter measures. In data replacement attack (DRA) the query issuing node will not get the global top-k query score earlier so the malicious node will replaces high-score value with its own low score value. The query issuing node will believe that the data it receive is a correct global top-k query result and it is difficult to detect the attack. DRAs attack are stronger than any other traditional attacks so some specified type of mechanism is need to overcome the DRA attack and more difficult to detect than other traditional types of attack, so some specific mechanism for defending DRAs are required. In this paper, we propose top-k query processing along with the method to detect DRA attack in MANETs. In the query processing to maintain accuracy of the top-k query score and detect the attacks, the replay data item contains results and also include the path information using which the query issuing node can determine the replay path and assure that results are from authentic nodes. The query issuing node can narrow down to the malicious node using the received path information and can request to send down the data item again. In this way, the

query-issuing node can identify the malicious node [1]. But in typical network there will be more number of malicious nodes are available in the network it is difficult to identify them using a single query message. But the proposed method is designed to identify the malicious node in an efficient way in which if a node find its nearby node to be an malicious one and this inform is shared with all other normal nodes with in the network in this way all the nodes will have global picture about the malicious nodes. In this case, sometime a normal node can send false information for this type of attack we proposed a new method that finds false notification attack (FNA). Each nodes in the network shares the information about the malicious node they have classified based on the path and data information with this information the normal nodes can identify the attackes even if the malicious nodes mixes the information as that of normal node

II. LITERATURE SURVEY

In MANET, secure routing protocols protect against attacks and false data. In these protocols data transmission from source to destination occurs in multiple routes [15], [8], [11] and public keys are symmetric keys are used for data encryption [6], [9], [13]. In [15], the authors proposed a method where every sensor nodes forwards data items using Message Authenticate Code (MAC). MAC uses symmetric key for encryption. Whenever the node receives a message, it checks the validity of message. Even if the information encrypted data replacement attack cannot be avoided. In [11], authors proposed a method in which multiple routes are determined. The route request messages are encrypted using hash functions. Top-k query is effectively used in the field of distributed and database systems to retrieve only the necessary data items from huge amount of data. In [1], [2], [4] and [10], authors proposed methods which adapts to mobility, provides high accuracy and reduces congestion. In [6], authors proposed secured query processing method in a network which contains malicious node. In [5], a method proposed to detect false data injection attack in which new and false data are generated by malicious node. In [3], [7], [12] and [14], methods for many reputation systems are proposed. In [19] and [20], each mobile node manages the neighboring nodes reputation values. By analyzing the messages of neighbor node, each node determines the reputation value. In [17] and [18], authors proposed a method in reputation system which is against the false notification attack. This method exchanges a cryptographic key between sender and receiver in advance. Also, sends their ID with past and present reputation scores in encrypted form. The receiver node can decode and confirm the received reputation scores. So that false reputation scores can discard n networks where aggregation is used to get the result on sensor network the main objective is to classify those node that are trusted node that contribute to the aggregation calculation. In [2] a secure hierarchical innetwork aggregation in used to identify the misbehave node and prevent them from participating in the aggregation calculation. In ad hoc networks [3], to obtain only the needed data items effectively each mobile node retrieves data items using a top-k query. In order to reduce the traffic with high accuracy of the query result, each node will send a histogram data [4] based on the query to query issuing node with this histogram it's easy to find the highest score value. A routing table based method is proposed in [5] to achieve high accuracy in query processing using top-k query. The top -k query processing is performed in two phase[6] in phase one the query issuing node will collect all the query result and in second phase it will find out the result with highest score based on threshold calculation. In two-tier sensor network [7] master slave architecture is used. In which the master node collects data from sensor node and answers the query from the network owner in this method the master node should be a trusted node. In co-operative Peer-to-peer (P2P) there is a possibility that one peer may cheat the other peer and propagate malicious code or some they don't cooperate [8]. The client server model of security in not sufficient for P2P so they use a cryptographic protocol along with the selfcertification In this segment, we survey existing studies on secure directing, top-k inquiry handling strategies, and notoriety frameworks.

A. Secure Routing Methods

In the field of MANET, secure directing conventions ensure against falsification of information and DoS attacks have been all around concentrated on. Secure directing conventions generally utilize information transmission along various courses (from the source hub to the destination hub) [6], [8], [9], and information encryption utilizing symmetric or open keys [6], [9], [7]. In [6], the creators have proposed a strategy in which the source hub decides numerous sheltered courses (from the source hub to the destination hub) by encoding the course ask message utilizing a hash capacity before sending information things.

B. Top-k Query Processing Methods

In the field of database frameworks and conveyed frameworks, top-k inquiry is viable to recover just the required information things in a extensive measure of information things. In [2], [5], [10], the creators have proposed strategies to decrease vitality consumption and traffic in unstructured P2P systems or remote sensor systems, by empowering hubs to lter pointless information things. In any case, these techniques don't ensure against DRA, also, are unsatisfactory for use in MANETs, since they are definitely not adjusted to hub portability. In our proposed method, the query issuing mobile node forwards the query to all other nodes in the network. When the neighbor node receives the query it stores the detailed information in all possible routes. Then k highest score values will be forwarded to two neighbor mobile nodes as a reply message. This information is also stored in forwarding routes which contains the sender node and receiver node IDs. This helps to detect an attack in the medium. In MANETs, there occurs a dynamic topology change due to mobility. When a link between two mobile nodes gets disconnected then the reply message is forwarded in an alternative route. The query issuing node narrows down

the malicious node after detecting DRA based on reply messages. The malicious nodes which are far can be identified by sharing the information of identified malicious node candidates. The nodes in the network can be divided into some groups based on the similarity of the received information

III. PROPOSED MODEL

The query-issuing node which issues a query firstly in the network sends a message for constructing routing table, and then nodes that receive the message reply the information on scores of data items held by them. The receiver stores the information in the received ranking table into own routing table. The receiver sets query addresses for all ranks in own routing table as the identifier of the node that sent the ranking table to it. The receiver updates query addresses for ranks to its own identifier if it holds the corresponding data items. In our proposed method, the query issuing mobile node forwards the query to all other nodes in the network. When the neighbor node receives the query it stores the detailed information in all possible routes. Then k highest score values will be forwarded to two neighbor mobile nodes as a reply message. This information is also stored in forwarding routes which contains the sender node and receiver node IDs. This helps to detect an attack in the medium. In MANETs, there occurs a dynamic topology change due to mobility. When a link between two mobile nodes gets disconnected then the reply message is forwarded in an alternative route. The query issuing node narrows down the malicious node after detecting DRA based on reply messages. The malicious nodes which are far can be identified by sharing the information of identified malicious node candidates. The nodes in the network can be divided into some groups based on the similarity of the received information.

IV. METHODOLOGY

A. *Network Creation*

The Network is constructed with 60 mobile nodes without any base station as self-distributed nodes. Each node is assigned with a unique identification number and mobility pattern is random. The node can exchange data packets and control packets as defined by the protocol.

B. *System Model*

The network consists of mobile node is represented by N= {N1,N2,…..Nn} where n is the total number of nodes in the network and they are identified using the identification number NID={NID1,NID2,..NIDm}, Where m=n. The data in the network is denoted as D={D1,D2,…Dk}, where k is the total number of data and each data is identified by using data identifier Di, where i=k. The algorithm works in distributed environment so each node has to exchange more information with the nearby node so they exchange data packet frequently so to avoid intermediate nodes not to modify the data content public key encryption method is used. Each node knows the public key of other nodes so data are send by encrypting with the public key of the receiving node. In order to reduce the computation the query message are not encrypted.

C. *Data Replacement Attack*

The node in the network can generate a query and send it to the all the nodes to get a desired value. Let us assume a node need the person detail with a particular blood group with high blood pressure, low vision, this requirement is generated as query and propagated towards the network. Let us consider Mr be the query issuing node and Mq be the node that replay for the query with its own score value this is the normal situation, the case will not remain for long time. In some situation a malicious node may capture the node and induce its own low score value to make the aggregation to be invalid. The query form query issuing node have a query id and the id of the query issuing node (Qid, Nid) the query goes to nearby node and this node will include its score value and its identification (SVi,, Nid).There will be two list one is to store the replay Score Value List (SVL) and the second is Replay Path Value (RPV) which store the path of the query propagation message or replay message. The query will take multiple path in the network. The query path is defined as the number of hop count which is calculated based on the network size and the radio range between nodes.

The waiting time for the replay is defined as the function of number of hop count between the source node and the replay node and the maximum size of the network along with the waiting time of query at each node. Replay_WT=(Net_Size – D_SR)* WaitTime Where Replay_WT is the replay waiting time for a query issuing node and Net_Size is the size of the network and D_SR the number of hops between the source node and current replay node, the waiting time is the time that a query takes to be processed at every node. The nodes will sends back the replay with its own identifier (Source node ID), and replay route (Dest node ID), a list of data items that containing the score values. The replay message includes two lists, Score Value List (Contains all the collected score values) and the second list is Replay Path Value (Contains all the nodes that as called forward nodes id).If the Replay Path Value has a node id but there is no date in the Score Value List this states that some replacement is done that is Data Replacement Attack is taken place.

The nodes will compare the score with the neighbor to detect the attacks.

Algorithm 1:
Detection of Attack
/* After the query-issuing node receives all reply messages */

1: INPUT: Top-k Result, Relpay_Message

2: OUTPUT: Route Information
3: Route Information ← ∅
4: for each Relpay_Message do
5: for each Top-k Result do
6: if Replay_path_Vale includes the node ID of a node processing a data item in Top-k Result and Score_Value_List does not include the data item then
7: Insert a route from the node with the missingdata item
8. the query-issuing node into SendRoute
9: end if
10: end for
11: end for
12: if Route Information= ∅ then
13: Detect Attack
14: end if

D. False Notification Attack

The nodes are grouped with some similar properties. Each group will have a group in-charge which is elected by Nodes highest ID. If some node inside any group identifies an attack based on the algorithm 1 it will report the malicious node id to is group in-charge and this information is shared with all other group in- charge inside the network. Each group in-charge will try to conform weather the node is malicious node or lire node (LN) .Where LN are normal node which will contribute a false value , No value to top-K query. The LN nodes will update the score value in Score_Value_List so it is not a malicious node, to verify this the query issuing node will send a request to this LN node to send its score value. Then the values are compared with the values collected from replay messages if the values are of in greater variation the LN nodes are categorized.

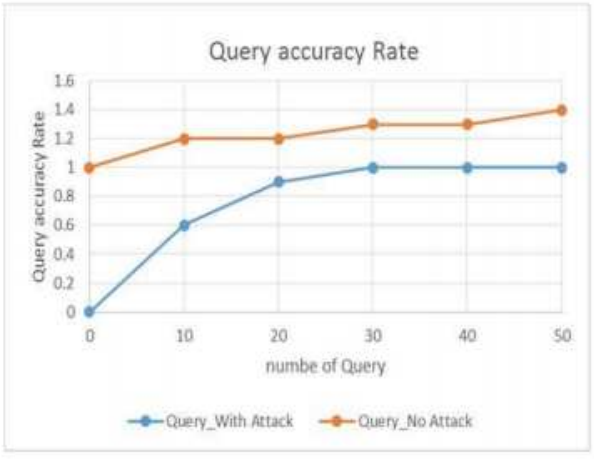

Figure 1. Query Result Accuracy

Figure.1 shows the accuracy of query result acquired by query-issuing node. The X-axis denotes the number of requested data items and Y-axis denotes the accuracy. The proposed top-k query method increases the accuracy even when the number of requested data items is large. Figure.2 shows the traffic occurred when query results are forwarded in multiple routes. The X-axis denotes the number of requested data items and Y-axis denotes the traffic. Figure.3 shows the malicious node identification ratio that represents maximum number of identified malicious node by issuing less number of queries. The X-axis denotes the query issuing time and misidentification.

The propose method shows the query result improves when the malicious nodes are identified and removed and also when the malicious nodes are present the query result accuracy is low as shown in figure 2. The figure 3 shows the traffic flow when the queries are issued inside the network and it is compared with the attack and without attack. The traffic is high when there is a malicious nodes inside the network since they contribute false information inside the network this lead the normal node to send more query to settle down on correct result.

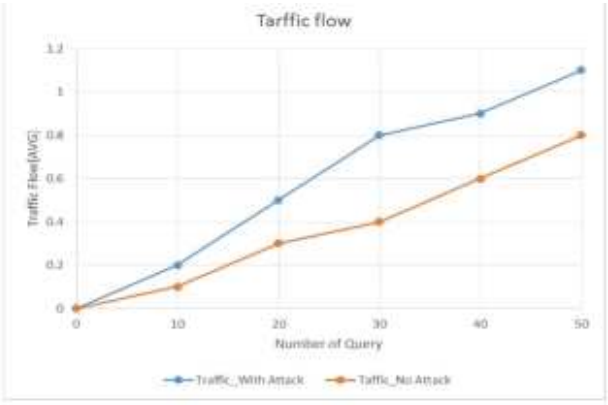

Figure 2. Traffic Flow Analysis

V. CONCLUSION

In this paper, we proposed node grouping methods for top-k query processing to identify multiple malicious node. To maintain high accuracy of reply message and to detect data replacement attack, k data items are transmitted along multiple routes. When query issuing node detects an attack it narrows down the malicious node candidates. Then malicious node identified by exchanging message with other nodes. Single query is not sufficient to identify multiple malicious nodes. So the information about identified malicious node shared with other nodes in network. In node grouping technique, nodes are divided into some group based on the similarity of the received information. Then, malicious nodes are identified based on group information. Since reply messages are transmitted along multiple routes, traffic in the network gets high. As a part of future work, a method can be proposed to reduce traffic and to provide message authentication.

REFERENCES

[1] R. Hagihara, M. Shinohara, T. Hara, and S. Nishio, A message processing method for top-k query for traffic reduction in ad hoc networks, in Proc. MDM, May 2009, pp. 11-20.
[2] D. Amagata, Y. Sasaki, T.Hara, and S.Nishio, A robust routing method for top-k queries processing in mobile ad hoc networks, in Proc. MDM, Jun. 2013, pp. 251-256.
[3] K. Liu, J. Deng, P. K. Varshney, and K. Balakrishnan, An acknowledgement-based approach for the detection of routing misbehavior in MANETs, IEEE Trans. Mobile Comput., vol. 6, no. 5, pp. 536- 550, May 2007.

[4] Y. Sasaki, T. Hara, and S. Nishio, Two-phase top-k query processing in mobile ad hoc networks, in Proc. NBiS, Sep. 2011, pp. 42-49.
[5] C.-M. Yu, G.-K. Ni, I.-Y. Chen, E. Gelenbe, and S.-Y. Kuo,Top-k query result completeness verification in tiered sensor networks, IEEE Trans. Inf. Forensics Security, vol. 9, no. 1, pp. 109-124, Jan. 2014.
[6] R. Zhang, J. Shi, Y. Liu, and Y. Zhang, Verifiable fine-grained top-k queries in tiered sensor networks, in Proc. INFOCOM, Mar.2010, pp. 1-9.
[7] 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.
[8] S. J. Lee and M. Gerla, Spilt multipath routing with maximally disjoint paths in ad hoc networks, in Proc. ICC, vol. 10. Jun. 2001, pp. 3201-3205.
[9] 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.
[10] Y. Sasaki, R. Hagihara, T. Hara, M. Shinohara, and S.Nishio, A top-k query method by estimating score distribution in mobile ad hoc networks, in Proc. DMWPC, Apr. 2010, pp. 944-949.
[11] B. Malhotra, M. A. Nascimento, and I. Nikolaidis, Exact top-k queries in wireless sensor networks, IEEE Trans. Knowl. Data Eng., vol. 23, no. 1 0,pp. 1513_1525, Oct. 2011.
[12] M. Wu, J. Xu, X. Tang, and W. C. Lee, Top-k monitoring in wire-less sensor networks, IEEE Trans. Knowl. Data Eng., vol. 19, no. 7,pp. 962_976, Jul. 2007.
[13] Y.-C. Hu, D. B. Johnson, and A. Perrig, SEAD: Secure efficient distance vector routing for mobile wireless ad hoc networks, Ad Hoc Netw., vol. 1,no. 1, pp. 175_192, Jul. 2003.
[14] X. Liu, J. Xu, and W. C. Lee, A cross pruning framework for top-kdata collection in wireless sensor networks, in Proc. MDM, May 2010,pp. 157_166.
[15] S. Kurosawa, H. Nakayama, N. Kato, A. Jamalipour, and Y. Nemoto,Detecting blackhole attack on AODV-based mobile ad hoc networks bydynamic learning method, Int. J. Netw. Secur., vol. 5, no. 3, pp. 338_346,2007.
[16] H. Chan, A. Perrig, and D. Song, Secure hierarchical in-network aggregation in sensor networks, in Proc. CCS, 2006, pp.278_287.
[17] S. Chen, Y. Zhang, Q. Liu and J. Feng, Dealing with dishonest recommendation: The trials in reputation management court, Ad Hoc Netw., vol. 10, no. 8, pp. 1603-1618, Nov.2012.
[18] P. Dewan and P. Dasgupta, P2P reputation management using distributed identities and decentralized recommendation chains, IEEE Trans. Knowl. Data Eng., vol. 22, no. 7, pp. 1000-1013, Jul. 2010.
[19] S. Buchegger and J.-Y. Le Boudec, Performance analysis of the CONFIDANT protocol, in Proc. MobiHoc, 2002, pp. 226-236.
[20] S. Marti, T. J. Giuli, K.Lai, and M. Baker, Mitigating routing misbehavior in mobile ad hoc networks, in Proc. MobiCom. 2000, pp. 225-265.