International Journal of Computer Science and Engineering Communications,
Vol.5, Issue.2 (2017):Page.1542-1549
IJCSEC.COM

An Efficient Identification of Malicious Node in Top-k Query Processing

T.Mohanapriya[1], S. Raja Ranganathan[2] and S. Karthik[3]

[1]PG Scholar, Dept. of CSE, SNS College of Technology, Coimbatore, India, Email: priyagvmp@gmail.com
[2]Assistant Professor, Dept. of CSE, SNS College of Technology, Coimbatore, India, Email: rajaranganathan@hotmail.com
[3]Dean, Dept. of CSE, SNS College of Technology, Coimbatore, India, Email: profskarthik@gmail.com

Abstract-*In mobile ad hoc networks (MANETs), data items can be retrieved effectively by using top-k 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.*

Keywords: *Mobile ad hoc networks, query processing, routing, traffic, data replacement attack, node grouping.*

I. Introduction

The mobile ad hoc network (MANET) is fewer infrastructures, decentralized and self configuring network. This is constructed by using the mobile nodes which are linked by wireless. Every individual mobile node in medium take action as a router and the nodes can be in touch with each other by inter-changing the information packets. Yet if initiator and target mobile nodes are not in the communication range, the information packets can be forwarded to the target node through the in-between nodes which exists among the two mobile nodes. Multiple hops are necessary for a node to communicate with further node across the network. The mobility in the network is high since the mobile nodes can move separately in any direction. Owing to this, the link gets detached and the network topology changes quickly. Because of these dynamic activities, more routing protocols like proactive, reactive and hybrid are anticipated for MANET. Each mobile node has poor communication bandwidth and limited battery life span for data transmission. This paper, discuss about the procedure of top-k query and the methodologies proposed for reducing traffic and providing high accuracy of query results. The traffic in the medium will be high when enormous amount of data packets are transmitted. To stay away from the unnecessary traffic, the node retrieves only the required data packets. For this reason, a technique called top-k query is proposed. It is efficient to retrieve only essential data items in the huge amount of data items. At this time, the data items are arranged by the scores of a particular attribute. A node which retrieves the data items is query-issuing node. This node floods a query note to every other node in the network. All mobile node transmits its data items with k maximum score (local top-k result) after getting the query message. The rest of this paper is prepared as follows: Section II deals with related works of top-k query. In Section III, we present the proposed methods for top-k query. In Section IV, Simulation experiments are presented. Finally, in Section V, we conclude this paper.

II. Related Works

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.

III. Proposed Model

A. Overview

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.

B. Top-k Query Processing

1. Query forwarding and replying

First, the query issuing nodes Mq floods query message to entire medium. It consists of query-issuing node identifier (Q-I_ID), query identifier (Q_ID), number of requested data (k), condition of query and the list of nodes in the path (Q_path).Mp sets waiting time (WT) in (1) for reply messages. In Algorithm Forwarding Query, hopCount denote number of hops to query issuing node.

WT= (maxhop - maxcount) . Twait (1)

When receiver node Mr receives query it stores the sender node ID and query path. In Algorithm Replying Query, when Mr needs to send reply message (RM), it selects the least hop count neighbor node from the information stored in forwarding list route (RM_FR).

Algorithm: Forwarding Query

```
If Mr receives query for first time then
  Store Q_path and hopcount as parent path
  Store nodeID in Q_path as parent
  Set WT for replying messages
  Send query to neighbor nodeID
Else
  Store Q_path and hopcount as neighbor path
  Store nodeID in Q-path as neighbor
End if
```

2. Link Disconnection

In MANETs, topology of network changes frequently due to high mobility. When a node Mr tries to forward a reply message to neighbor node and link gets disconnected, it results in decreasing the accuracy of query result. To overcome this problem, whenever a node sends reply message it waits for the acknowledgement ACK from the sender node. When Mr doesnot receives ACK from parent node it detects the disconnection in radio link. Then, Mr sends data items through another neighbor node which has least hop count.

Algorithm: Replying Query

```
For each neighbor do
  If hopCount of neighbor is minimum then
      Insert Neighbor as Destination
  End if
End for
Add local result to RM
For i=0 to 1 do
  If i=0 then
      Add ( Mr,parent ID) to RM_FR and send RM to parent ID
  Else if i=1 then
      Add (Mr,Destination) to RM_FR and send RM to Destination
  End if
End for
```

3. Attack Detection

After receiving all reply messages, the query issuing node Mp detects DRA. In Algorithm Detection of Attack, T-k result denotes highest scores k. RM_Data and RM_FR denotes data list and forwarding route respectively. Sendroute denotes set of node identifiers. A node can detect DRA when data items in T_k included in Sendroute but not included in RM_Data. To identify malicious node, query issuing node narrows down the candidates of malicious node in Sendroute. Miss T_k result denotes replaced data. If number of candidate is one, query node identifies this as malicious node. If number of candidates is more than one, an inquiry message M_INQ sends to other nodes. By reply message for inquiry query the malicious node candidate can be identified.

```
Algorithm: Detection of Attack
If hop count to candidate = 1 then
  Return candidate as malicious node
Else if hop count to candidate is > 1 then
  Send M_INQ to Mdes
End if
If Mdes receives M_INQ then
  Send M_Rep send by Candidate[i] to Mp
End if
If Mp receives M_REP then
  If scores includes in scores of Miss T_k result then
Return candidate [i-1]
  End if
End if
```

4. Global Identification

Each mobile node forms groups in medium based on the conventional information in notification messages. In Algorithm Node Grouping, sim(x,y) denotes resemblance of scores between Mx and My. Grp and G_CAN represent groups and candidates of groups respectively. Grp_f denotes fth group of Grp. M_h denotes node in Grp_f. First, each mobile node calculates similarity of malicious nodes based on received messages. Cosine similarity made in order to decrease the power of differences in recognized malicious node. After node grouping, a few groups may include both normal and malicious nodes. As a result, node performs cleaning in every group to eliminate contradiction. Following that, node identifying one more node which is identified by less than a certain integer of nodes in same group, is also removed from group.

```
Algorithm: Node Grouping
For each x ϵ n do
  For each y ϵ n do
sim(x,y) = cos(x,y) = $Rx.Ry/|Rx||Ry|$
  End for
End for
For each x ϵ n do
  For each y ϵ n do
    If M_CAN = ø and sim(x,y) >= ø then
insert Mx, My into G_Can
    Else if M_CAN ≠ ø and { $\forall z \in$ M_CAN, sim(z,y) ≥ $\theta$ } then
Insert Mx into G_Can
    End if
  End for
  For each Grp do
    For each $M_h$ in $Grp_f$ do
      If $M_h$ identifies node include in $Grp_f$ then
Eliminate $M_h$ from $Grp_f$
      End if
   End for
  End for
End for
```

IV. Simulation Experiment

This section deals with the results of simulation experiments conducted using the network simulator NS2. By using random waypoint model nodes are created and initial position determined randomly. The data items from each mobile node are transmitted using IEEE 802.11b device.

Performance Metrics:

The following three performance metrics are measured in this simulation.

1. Accuracy of query result: It represents the average ratio of data items provided in the top-k result acquired by query-issuing node.
2. Traffic: It represents the total volume of traffic taken for processing the query and for detecting the malicious node.
3. Malicious Node Identification Ratio: It represents the average ratio of the number of identified malicious node using node grouping technique.

Simulation Results:

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.

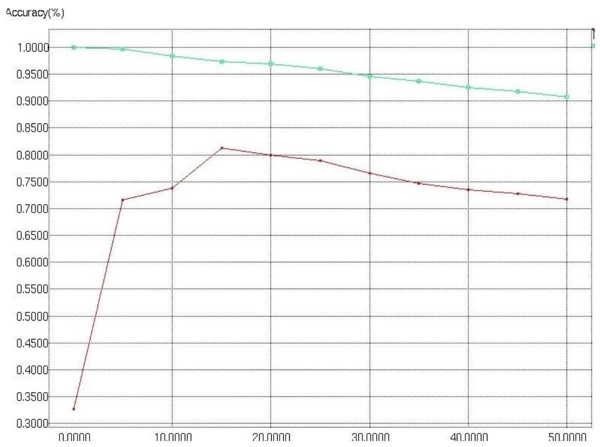

Figure 1: Accuracy of Query Result

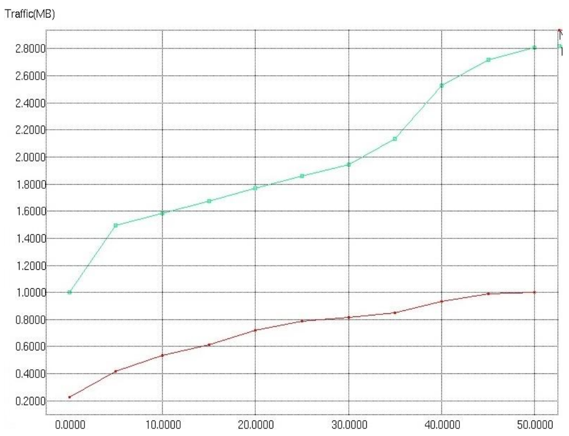

Figure 2: Traffic

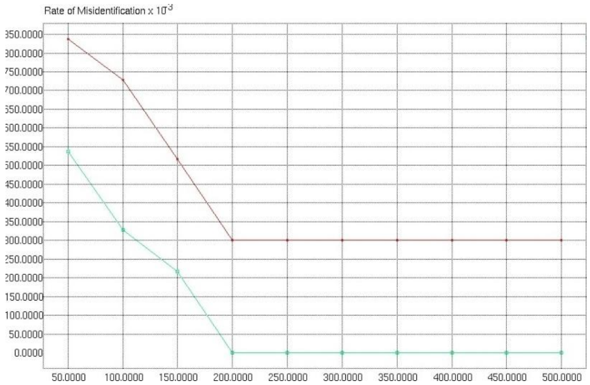

Figure 3: Malicious node identification ratio

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. 10,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.