International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 10
ISSN: 2454-4248
14 – 19

Identification of Malicious node for Effective Top-k Query Processing in MANETS

H. Swathi
PG Scholar,
Department of CSE,
MVSR Engineering College.

Dr. Akhil Khare
Head of the Department,
Department of CSE,
MVSR Engineering College.

Abstract- In Mobile Ad-hoc networks, query processing is optimized using Top-k query processing. The accuracy of the results can be lowered if there exists malicious node. In our proposed system, we assume that malicious node perform Data Replacement Attack, in which the malicious node replaces necessary data sets with the false data sets. In our system malicious node identification method, the query issuing node receives the reply messages from the nodes; if a query-issuing node detects a DRA then it performs subsequent inquiries with the nodes which receive the information from the malicious node. In this way the query issuing node identifies the malicious node, and shares the information with the neighbouring nodes. Then the nodes share the information regarding the malicious node with the other nodes which are far away. Each node tends to identify the malicious node in the network, and then floods the information. Query issuing node performs grouping of the nodes based on the similarity of the information on malicious node detected by the nodes. Identification of malicious node is performed based on the results of malicious node identifications by these groups.

Keywords: *Ad hoc networks, top-k query processing, data replacement attack, grouping.*

*****

I. Introduction

An ad-hoc network is a collection of wireless mobile hosts forming a temporary network without the aid of any stand-alone infrastructure or centralized administration. Mobile Ad-hoc networks are self-organizing and self re-configuring multi-hop wireless networks where, the structure of the network changes dynamically. This is mainly due to the mobility of the nodes. Nodes in these networks utilize the same random access wireless channel, cooperating in a friendly manner to engaging themselves in multi-hop forwarding. The nodes in the network not only act as hosts but also as routers that route data to/from other nodes in network. In mobile ad-hoc networks where there is no infrastructure support as is the case with wireless networks, and since a destination node might be out of range of a source node transmitting packets; a routing procedure is always needed to find a path so as to forward the packets appropriately between the source and the destination.

MANETS rely on wireless transmission, a secured way of message transmission is important to protect the privacy of the data. An insecure ad-hoc network at the edge of an existing communication infrastructure may potentially cause the entire network to become vulnerable to security breaches. In mobile ad hoc networks, there is no central administration to take care of detection and prevention of anomalies. Mobile devices identities or their intentions cannot be predetermined or verified. Therefore nodes have to cooperate for the integrity of the operation of the network. However, nodes may refuse to cooperate by not forwarding packets for others for several reasons and not want to exhaust their resources. Various other factors make the task of secure communication in ad hoc wireless networks difficult include the mobility of the nodes, a promiscuous mode of operation, limited processing power, and limited availability of resources such as battery power, bandwidth and memory. Therefore nodes have to cooperate for the integrity of the operation of the network. Nodes may refuse to cooperate by not forwarding packets for others for selfish reasons and not want to exhaust their resources.

In ad hoc networks devices (also called nodes) act both as computers and routers. Most routing protocols lead nodes to exchange network topology information in order to establish communication routes. This information is sensitive and may become a target for malicious adversaries who intend to attack the network or the applications running on it. There are two sources of threats to routing protocols. The first comes from external attackers. By injecting erroneous routing information, replaying old routing information, or distorting routing information, an attacker could successfully partition a network or introduce a traffic overload by causing retransmission and inefficient routing. The second and more severe kind of threat comes from compromised nodes, which might (i) misuse routing information to other nodes or (ii) act on applicative data in order to induce service failures. The provision of systematic approaches to evaluate the impact of such threats on particular routing protocols remains an open challenge today.

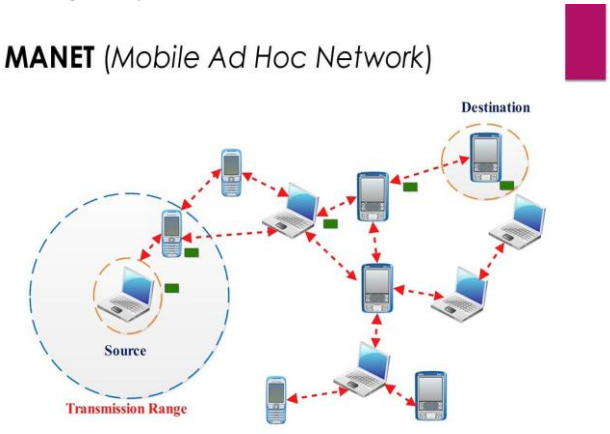

Figure: Mobile Ad Hoc Network model

 ISSN: 2454-4248
14 – 19

MANET Characteristics

In MANET, each node act as both host and router. That is it is autonomous in behavior. Multi-hop radio relaying- When a source node and destination node for a message is out of the radio range, the MANETs are capable of multi-hop routing. Distributed nature of operation for security, routing and host configuration. A centralized firewall is absent here. The nodes can join or leave the network anytime, making the network topology dynamic in nature. Mobile nodes are characterized with less memory, power and light weight features. The reliability, efficiency, stability and capacity of wireless links are often inferior when compared with wired links. This shows the fluctuating link bandwidth of wireless links. Mobile and spontaneous behavior which demands minimum human intervention to configure the network. All nodes have identical features with similar responsibilities and capabilities and hence it forms a completely symmetric environment. High user density and large level of user mobility.

II. Related Work

"A Routing Method for Top-k Query Processing in Mobile Ad Hoc Networks", Daichi Amagata, Yuya Sasaki, Takahiro Hara, Shojiro Nishio Recently, mobile ad hoc networks (MANETs) have been attracting increasing interests. Query processing for retrieving data items is one of the important issues in MANETs. Since the network bandwidth and batteries of mobile nodes are limited in MANETs, it is important for query processing to acquire only necessary data items in order to reduce data traffic. A possible and promising solution is that each mobile node acquires data items using a top-k query, in which data items are ordered by the score of a particular attribute and the query-issuing node acquires data items with the k-highest scores (top-k result). A naive manner to process a top-k query in a MANET is as follows. A query-issuing node floods all mobile nodes in the entire network with a query message, and then, each mobile node that receives the query message transmits its own data items with k-highest scores. By doing so, the query-issuing node can acquire all data items included in the top-k result if communication failures do not occur. However, many data items that are not included in the top-k result are sent back to the query-issuing node. This causes communication failures and large battery consumption.

"A Self-Organized Mechanism for Thwarting Malicious Access in Ad Hoc Networks" Natalia Castro Fernandes, Marcelo Duffles Donato Moreira, and Otto Carlos Muniz Bandeira Duarte

Ad hoc networks do not rely on a physical infrastructure or a central administration entity. Indeed, a different user controls each node and therefore security becomes a major issue to keep collaborative message forwarding working. To restrict undistinguished node access in regular networks, two complementary approaches are used: access control and authentication. When we can authenticate nodes and identify malicious actions, the network is able to punish malicious nodes and reward the cooperative ones. In ad hoc networks, however, both access control and authentication are challenging, because they are usually based on centralized mechanisms. Accordingly, ad-hoc networks demand self-organized mechanisms based on distributed administration and nodes with equivalent functions, providing high availability even on network partitions.

"Progressive Distributed Top-k Retrieval in Peer-to-Peer Networks" Wolf-Tilo Balke, Wolfgang Nejdl Wolf Siberski, Uwe Thaden With information needs emerging beyond a simple exact match paradigm, databases and information systems have since long catered for extended retrieval paradigms like top-k retrieval or skyline queries. Query languages like SQL over relational databases have been extended to facilitate rank- and/or score-based retrieval algorithms assigning a degree of match with respect to all query predicates to each database object and then aggregating the rank/score values to get only the set of best matching answers. Moreover, the new retrieval paradigms allow for the direct incorporation of user preferences into queries for a more cooperative retrieval behaviour. Whereas too specific query predicates under the exact match paradigm would far too often lead to empty result sets, the notion of best matches and relative importance of predicates can thoroughly satisfy a user's information needs independent of the respective database instance. Especially top-k queries delivering a well defined set of k best answers according to a user-provided, probably weighted compensation function have shown their broad applicability in various areas like Web search engines, mobile database applications, or content-based retrieval in multimedia collections or digital libraries.

"An Acknowledgment-Based Approach for the Detection of Routing Misbehavior in MANETs" Kejun Liu, Jing Deng, Pramod K. Varshney, and Kashyap Balakrishnan A Mobile Ad Hoc Network (MANET) is a collection of mobile nodes (hosts) which communicate with each other via wireless links either directly or relying on other nodes as routers. The operation of MANETs does not depend on pre-existing infrastructure or base stations. Network nodes in MANETs are free to move randomly. Therefore, the network topology of a MANET may change rapidly and unpredictably. All network activities, such as discovering the topology and delivering data packets, have to be executed by the nodes themselves, either individually or collectively. Depending on its application, the structure of a MANET may vary from a small, static network that is highly power-constrained to a large-scale, mobile, highly dynamic network.

"Ariadne: A Secure On-Demand Routing Protocol for Ad Hoc Networks" Yih-Chun Hu, Adrian Perrig, David B. Johnson An ad hoc network is a group of wireless mobile computers (or nodes), in which nodes cooperate by forwarding packets for each other to allow them to communicate beyond direct wireless transmission range. Ad hoc networks require no centralized administration or fixed network infrastructure such as base stations or access points, and can be quickly and inexpensively set up as needed. They can be used in scenarios in which no infrastructure exists, or in which the existing infrastructure does not meet application requirements for reasons such as security or cost. Applications such as military exercises, disaster relief, and mine site operation may benefit from ad hoc networking, but secure and reliable communication is a necessary prerequisite for such applications.

III. Problem Statement

While performing Top-k query processing, there may exist a malicious nodein the network. This malicious node lowers accuracy of the query result, performingData Replacement Attack.

 ISSN: 2454-4248
14 – 19

Data Replacement Attack, in which a malicious node replaces the received data items with false yet proper data items. DRA is a strong attack, and more difficult detect.So, Identification of malicious node for effective top-k query processing in MANETs, method is used in order to identify the malicious node present in the network.

IV. Proposed System Implementation

In the proposed method to identify the malicious node the system environment is assumed to be a MANET. We assume a set of mobile nodes in the system, and assign one particular node as a Query issuing node. All the other nodes in the network sends reply messages to the query issuing node. Query issuing node receives the reply messages from other nodes. When a node sends a the reply message it encrypts the data items using the public key of the destination node to avoid other nodes overhearing the reply messages. But when the query issuing node floods the query in the entire network then the query is not encrypted because the query is not sent to a particular destination node but is sent to all the neighboring nodes. Also the scores in the data items are not encrypted because every node calculates the top-k highest result comparing its own scores with the other neighboring node and sends the result to the two other neighboring nodes and the process continues.

In the proposed method we assume that there exists a malicious node in the network, which performs the data replacement attack. When the query issuing node floods the query in the network and then when all the nodes reply with the data items and when the malicious node receives the data items from its neighbouring node then it replaces the highest scores with the lower scores. As the query issuing node receives the data items from multiple paths, when it receives the data items from the path in which the malicious node is present it gets aware of the data items replaced. Because the highest scores received from the other path differs from that of the path in which the malicious node exist. And then after the subsequent inquiries with the other nodes it determines that a particular node is a malicious node. Then it sends the notification that there is a malicious node in the network, and then all the other nodes perform grouping among themselves to identify the malicious node globally. Each node individually identifies the malicious node and then based on the majority of the nodes which concluded a particular node as a malicious node, then that particular node is finally identified as a malicious node in the network.

V. Proposed Method to Identify the Malicious node

In our proposed method, where top-k query processing is performed there is a query issuing node which floods the query i.e., to reply with the highest scores of the nodes. In the network there exists a malicious node which performs data replacement attack. In order to identify the malicious node, query issuing node and the normal nodes which are not malicious performs the following operations in the network. The steps to identify the malicious node are as follows:

Forwarding the Query

Let us assume a node as a Query issuing node, N_q . N_q floods the query over the entire network which includes query condition and the identifier of query issuing node(N_q), identifier of the query, and the list of the path to which query should be forwarded(Query Path). Query issuing node specifies the query to return the k data items to it's neighboring node N_p. nodecount denotes the number of nodes in between the nodes present in the path to the query issuing node. N_p sets the waiting time(W_t) to receive the reply from the nodes to which it has sent the query, based on nodecount and the $nodecount_{max}$.

$$W_t = (nodecount_{max} – nodecount)$$

Algorithm Used

1. if $N_p$receives the query for the first time
2. then store Querypath and the nodecount as parent query path
3. store the Queryissuing node ID as parent node ID
4. Set the W_t for receiving reply from the neighboring nodes
5. Forward the query to the neighboring node by adding N_p's node ID to the end of Querypath
6. else
7. store nodecount and Querypath as it's neighbour Querypath
8. store the nodeID as it's neighbour Querypath
9. end if

Sending the reply message

When W_t has elapsed, each node sends back the reply message from which it has received the query. The reply message includes its own node identifier, and the destination node's ID to which it has to send the reply message, list of data items which contains the scores of the respective nodes and the forwarding route list which contains the sender nodes and the destination nodes.

Let N_r denotes the node to which N_p has sent the query. Then N_r sends the reply message to N_p once the W_t has elapsed. N_r sends the reply message(REP) and the forwarding route list(REP.FRL) which contains the sender nodes and the destination nodes list. And N_r then sends the query to one of it's neighboring node which has minimum node count.

Algorithm Used

1. for each neighbor do
2. select a neighbor with minimum hop count
3. end for
4. add the top-k result to the forwarding route list(REP.FRL)
5. if
6. node is parent node send the result to the parent node.
7. else if
8. node is the neighboring node send the result to the neighboring destination node.
9. end if
10. end for
11. acknowledge the sender node of the reply
12. if
13. the node receives the top-k highest result than the other k-top scores which it has already received then
14. Send the REP which includes the top-k result to the parent node and destination node.
15. end if
16. if

17. N_r doesnot get the acknowledgement even after W_t then resend the REP to the destination node or the parent node.
18. end if

Detecting attack

Query issuing node, N_q detects the data replacement attack after receiving the reply messages(REP) which includes Data Items(REP.DI) with k highest scores and Forwarding route list(REP.FRL) (which includes the sender node ID and the destination node ID to which the node has sent the data) from all the nodes. If the node which possess the top-k score is included in the REP.FRL list but the score is not included in the REP.DI then insert the route from that node to the Query issuing node and denote it as (REPLYROUTE) then the query issuing node detects that the Data Replacement Attack. After detecting the attack the query issuing node initiates the malicious node identification.

Algorithm Used

1. for each REP
2. for each top-k result
3. if REP.FRL contains the node which possess the top-k result and the top-k result is not included in REP.DI then
4. add the route into REPLYROUTE from the node for which the data item is missing to the Query issuing node
5. end if
6. end for
7. end for

Sorting of the malicious node candidates

If the nodes are included in the REPLYROUTE then they are termed as suspected nodes and are included into SUSPECTEDNODES list by the query issuing node.

Algorithm Used

1. for each
2. node in REPLYROUTE do
3. if node is included in REPLYROUTE then
4. include the nodes into SUSPECTEDNODES
5. And malicious node identification is performed.

Malicious node identification

After the query issuing node gets the information regarding the suspected nodes then it performs the malicious node identification. Query issuing node starts the inquiry with the nodes to which the suspected node has sent the reply message. INQ denotes the inquiry message which includes the node ID of the destination node ID ($N_{d)}$(this is the node to which the suspected node has sent the data items) and query issuing node ID and also the route of the inquiry which is the route from query issuing node to the destination node (ROUTEINQ). REPINQ denotes the reply for the inquiry message from the destination node(N_d).

Algorithm Used

1. for each i insuspected nodes.sizedo
2. if suspected node>1then
3. send the INQ to the destination node (N_d)to which the suspected node has sent the data items
4. end if
5. if
6. a node which is not present in SUSPECTEDNODE list receives the INQ then the node forwards the INQ message to the other node.
7. end if
8. if
9. the N_d has received the INQ
10. send the REPINQ which includes the data items sent by suspected node[i] to N_d to the query issuing node.
11. end if
12. if
13. N_q(query issuing node) receives the REPINQ
14. if
15. the scores include the score of missing top-k result
16. then
17. return Suspectednode[i-1]
18. end if
19. end if
20. end for

Sharing information regarding the presence of malicious node

The query issuing node shares the information regarding the malicious nodes present in the network. The notification message includes the identified malicious nodes.

Identification of malicious node individually by each node

In this process each node behaves as a query issuing node. Identification of malicious node is performed by Node Grouping. All the nodes acts as query issuing nodes and sends the queries to the other neighboring nodes, and then receives the replies from the corresponding neighboring nodes. All the nodes replies with the notification messages which includes the identified malicious nodes done by themselves according to the data items sent and received from the other nodes.

VI. Node Grouping

Each node divides nodes in the network into some groups based on the information in the notification messages received by the other nodes. Each node calculates the similarity(θ) of nodes based on the identification of the malicious nodes. Based on the notification messages if two nodes identify same nodes as malicious nodes then the similarity is 1. And if the two nodes identify the different nodes as malicious nodes then the similarity is set to 0. If only some of the nodes are matched in the identification of malicious nodes then θ is set to 0.7.

Query Issuing Node	Identified Malicious Node
N1	N2, N5
N2	N1, N4
N4	N2, N5
N6	N5
N5	N1, N4
N3	N5
N8	N1,N2,N5
N7	N5
N9	N5, N8
N10	N5

 ISSN: 2454-4248
14 – 19

In the above images all the nodes acts as query issuing nodes and identifies some of the nodes as malicious nodes. The groups are formed as follows when similarity=0.7

G_n denotes the groups, G_1={N_1,N_3,N_4,N_6,N_7,N_{10}}, G_2={N_1,N_4,N_8}, G_3={N_2,N_5},G_4={N_3,N_6,N_7,N_9,N_{10}}. In group 2 N_8 terms N_1 as malicious node, this should not be done because both belongs to the same group. So N_8 is eliminated from G_2. Now $G_{2=}$={N_1,N_8}. In the above table only N_9 identifies N_8 as malicious node, hence M_9 is also eliminated. Now G_4={N_3,N_6,N_7,N_{10}}. Final malicious node identification is performed after the node grouping is done. There are three types of groups, (i). normal nodes, (ii). malicious nodes, (iii). Normal and malicious nodes. The nodes which are identified as malicious nodes from the groups (i) and (ii)are malicious nodes. But in group (iii) there are normal and malicious nodes, and so normal nodes can be identified as malicious nodes. As, the malicious nodes are minority in the group and normal nodes are majority in the group. On the majority based conclusion, a particular node is identified as malicious node based on the majority of the nodes belonging to various groups.

Algorithm Used

1. for each node N_i
2. foreach node N_j
3. calculate the similarity(θ) of identification of malicious nodes by the nodes
4. if θ=1 between the nodes N_i and N_j then
5. N_i and N_j belongs to the same group
6. if θ>=0.7 between the nodes N_i and N_j then
7. N_i and N_j belongs to the same group
8. if θ>=0 between the nodes N_i and N_j then
9. N_i and N_j does not belongs to the same group
10. end if
11. end if
12. end if
13. end for
14. end for

VII. Final Identification of Malicious node based on Majority Method

Now the final malicious node is identified based on the number of nodes which belongs to the different groups have declared some of the nodes as Malicious nodes. In the above table nodes belonging to the different groups have declared some of the nodes as malicious nodes. As N_5 is identified as malicious node by the majority of the groups, hence N_5 is declared as malicious node in the network.

VIII. Result analysis

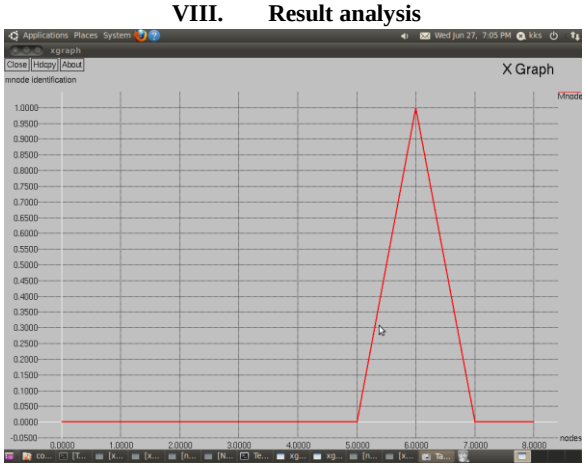

Figure: Malicious node detection

Group	Nodes in the group	Malicious Nodes
G1	N1, N3, N4, N6, N7, N10	N5
G2	N1, N4	N2, N5
G3	N2, N5	N1, N4
G4	N3, N6, N7, N10	N5

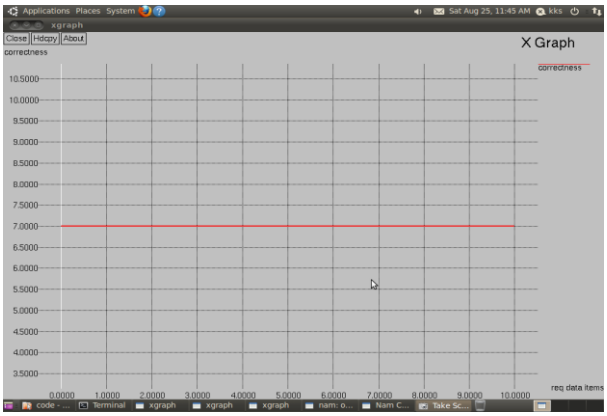

Figure: Accuracy of the data items

IX. Conclusion

In this system, we have proposed methods for identification of malicious node for effective top-k query processing.The detection of DRA attack is performed. This system helps in preventing or avoiding an attack in its initial stage. It can identify all the addresses of nodes in the selected routing path from a source to destination after the source has received the RREP message. This system helps in improving packet delivery rate. This system helps in achieving the reduced overhead.

As the future work we plan to implement the system for multiple malicious nodes and also design a message authentication method to prevent malicious nodes from performing false notification attacks.

 ISSN: 2454-4248
14 – 19

X. References

[1]. 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.

[2]. 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.

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

[4]. 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, pp. 483–502, Sep. 2002.

[5]. 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.

[6]. H. Chan, A. Perrig, and D. Song, ‘‘Secure hierarchical in-network aggregation in sensor networks,’’ in Proc. CCS, 2006, pp. 278–287.

[7]. 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.

[8]. P.DewanandP.Dasgupta,‘‘P2P reputation management using distributed identities and decentralized recommendation chains,’’IEEETrans.Knowl. Data Eng., vol. 22, no. 7, pp. 1000–1013, Jul. 2010.

[9]. N. C. Fernandes, M. D. D. Moreira, and O. C. M. B. Duarte, ‘‘A selforganizedmechanismforthwartingmaliciousaccessinadhocnetworks,’’ in Proc. INFOCOM, 2010, pp. 266–270.

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

[11]. Y.-C. Hu, A. Perrig, and D. B. Johnson, ‘‘Ariadne: A secure on-demand routing protocol for ad hoc networks,’’ in Proc. MobiCom, 2002, pp. 12–23.

[12]. Y.-C. Hu, D. B. Johnson, and A. Perrig, ‘‘SEAD: Secure efficient distance vectorroutingformobilewirelessadhocnetworks,’’AdHocNetw.,vol.1, no. 1, pp. 175–192, Jul. 2003.

[13]. S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, ‘‘The Eigentrust algorithm for reputation management in P2P networks,’’ in Proc. WWW, 2003, pp. 640–651.

[14]. K. Liu, J. Deng, P. K. Varshney, and K. Balakrishnan, ‘‘An acknowledgment-based approach for the detection of routing misbehavior in MANETs,’’ IEEE Trans. Mobile Comput., vol. 6, no. 5, pp. 536–550, May 2007.

[15]. S. Kurosawa, H. Nakayama, N. Kato, A. Jamalipour, and Y. Nemoto, ‘‘Detecting blackhole attack on AODV-based mobile ad hoc networks by dynamic learning method,’’ Int. J. Netw. Secur., vol. 5, no. 3, pp. 338–346, 2007.