A Survey on Traffic Pattern Discovery in Mobile Ad hoc Network

1 Priyanka Sen, 2 Vaishali Sahare

1 Department of Computer Science and Engineering, RTMNU,
G.H Raisoni Institute of Engineering and Technology for Women
Nagpur, Maharashtra, India

2 Department of Computer Science and Engineering, RTMNU,
G.H Raisoni Institute of Engineering and Technology for Women
Nagpur, Maharashtra, India

Abstract - Mobile Ad-hoc Network (MANET) is type of unguided network whose nodes are self-configuring and connected by wireless links. Anonymity communication is one of the major issues in MANET. Many anonymity enhancing techniques have been introduced based on packet encryption. These were designed for the protection of communication anonymity in mobile ad hoc networks. However, passive statistical traffic analysis attacks that can be vulnerable to MANET. The communication anonymity consists of two aspects: source/destination anonymity and end-to-end anonymity. In order to discover the traffic pattern without decrypting the captured packets, this proposed system will be designed. The proposed system will first search the required node by using a heuristic approach. Then statistical traffic analysis will be performed to find the data transmission of the searched node to its neighboring nodes. After performing the statistical traffic analysis, the probability of the searched node being a source or destination will be determined.

Keywords - Anonymity Communication, Mobile Ad hoc Network (MANET), Statistical Traffic Analysis.

1. Introduction

Mobile ad hoc network (MANET) is a self configuring wireless ad-hoc network of mobile devices. Every device in MANET is independent to move. This results in changes of link of such device. Thus MANET is also known as an infrastructure less network. The devices that are present in the network must forward the traffic to other devices. In MANET each device must act as a router. The basic snapshot of a mobile ad hoc network can be illustrated in Fig. 1. Each device is having a range associated with it. The devices can communicate with other devices within the range. The communication from one device to other may include other devices as intermediates. One of the critical issues in MANETs is communication anonymity. Anonymity can be defined as the state in which identity of an object that performs the action is hidden. An anonymous communication system can be defined as a technology that hides the object identity. In network communication anonymity refers to block out the identicalness of the nodes which are taking part in communication. The purpose of including the anonymous ways of communication ensures the concealment of the network from attacker. Communication anonymity [1] has two aspects: Source/destination anonymity and End-to-End relationship anonymity. In source/destination anonymity it is difficult to identify the sources or the destinations of the network flows while in end-to-end relationship anonymity it is difficult to identify the end-to-end communication relations.

In MANET communication anonymity has been proposed by anonymous routing protocols such as ANODR (ANonymous On-Demand Routing) [7], OLAR (On-demand Lightweight Anonymous Routing) [6]. All these anonymous routing protocols rely on packet encryption to hide the information from attackers. Still the passive attackers can eavesdrop on the wireless channel, intercept the transmission, and then perform traffic analysis attacks. Traffic analysis [10] is one of the types of passive attack in MANET. Traffic analysis is further subdivided into predecessor attack [8] and disclosure attack [9]. The following are the three nature of MANET due to which above approaches do not work well to analyze traffic in MANET:

1) Broadcasting nature: In wired networks point to point transmission can be easily applied to only one possible receiver. While in wireless network message is broadcasted to multiple receivers.
2) Ad hoc nature: Mobile node can be served as both source and destination. This can create confusion to determine the role of the node.

3) Mobile nature: Traffic analysis model do not consider the mobility of communication peers. This makes the communication among mobile node more complex.

There is a need of such a technology which can analyze traffic without any interruption of the above three characteristics of MANET. This proposed system fulfills the need. The objective of this paper is to show that passive attackers can perform traffic analysis without the knowledge of the adversaries. This approach is required in military environment. The proposed system will perform statistical traffic analysis to discover the traffic pattern. This system will perform the point to point as well as end-to-end traffic analysis among receivers. Indirectly this calculation will provide probable source and destination of the network that will discover the hidden traffic pattern. Thus the adversaries will not be able to know about the traffic analysis.

The remaining paper is organized as follows: Section 2 describes the related work. Section 3 presents the proposed system. Section 4 describes the proposed solution of the proposed system. Lastly section 5 presents the conclusion.

2. Related Work

Yang Qin, Dijiang Huang and Bing Li [1], proposed that though there are many anonymous routing protocols and anonymous enhancing techniques available still mobile ad hoc network (MANET) is vulnerable to passive statistical traffic analysis attacks. The authors proposed a system called as Statistical Traffic Pattern Discovery System (STARS). A STAR is used to discover the hidden traffic pattern in MANET. The drawback of this proposed system is that no searching algorithm is applied to search the traffic free path.

Douglas Kelly, Richard Raines, Rusty Baldwin, Michael Grimaila, and Barry Mullins [2], investigated on anonymity. For a user anonymity can be defined as using any services while keeping their identity hidden from an adversary. Anonymity help user to protect their data from attacks. Unidentifiability, Unlinkability, and Unobservability are the three properties of anonymity. Unidentifiability means the adversary is unable to determine one’s identity or action among similar ones. Unlinkability means the adversary is unable to relate messages or actions by observing the system. Unobservability means the adversary is unable to observe the presence of messages or action in the system. Since unobservability keeps the identity of messages or action secret it can be implied as anonymity. Unidentifiability is subdivided into sender anonymity (SA), receiver anonymity (RA), mutual anonymity (MA) and group anonymity (GA). Unlinkability is subdivided into location anonymity (LA), communication anonymity (CA) and group communication anonymity (GCA). In order to discover the traffic pattern we have to work on unidentifiability property of anonymity and decrease the sender anonymity (SA) and receiver anonymity (RA).

Lei Liu, Xiaolong Jin, Geyong Min, and Li Xu, [3] proposed that in order to detect the attack in a network traffic intensity and packet number are the two important metrics. Lei et al. had designed an anomaly detection system. This anomaly detection system is used to detect the distributed denial of service (DDoS) attack in MANET. When traffic analysis is carried on MANET these two metrics are used to detect the DDoS attack. Similarly when traffic analysis will be carried on our proposed system data transmission will be considered as a parameter. We can conclude that data transmission will be an important factor whenever traffic analysis will be carried out, though the reason may be for detection of attack or for discovery of traffic pattern.

Zhilin Zhang and Yu Zhang [4], introduced that control traffic plays an important role in route discovery in MANET. The characteristic that involve to carry out research on control traffic in MANET when on demand routing protocols are used include distribution of nodes’ control packet traffic, communication of control packets between nodes, rate of RREQ (route request) packets and the ratio of number of RREQ packets originating from one node to all RREQ packets relayed by this node. These characteristics of control traffic are affected by factors such as mobility, node density and data traffic. Thus theoretically we can determine that indirectly one of the
factors of control traffic is data traffic. Hence we can conclude that control traffic will also play an important role in route discovery in our proposed system though the situation will be different. In our proposed system we will find traffic free path i.e. control traffic path so that it will be easy to find out the number of data packets transmitted to neighboring nodes. This will help us to discover route in our proposed system.

Y. Liu, R. Zhang, J. Shi, and Y. Zhang [5] designed a novel algorithm called as traffic inference algorithm (TIA) which allows an adversary to infer the traffic pattern in MANET. This algorithm is based on the assumption that difference between data frames, routing frames and MAC control frames is visible to passive adversaries. Through these differences they can identify the point-to-point traffic using the MAC control frames, recognize the end-to-end traffic by tracing the routing frames and then find out the actual traffic pattern using the data frames. This algorithm is not a successful invention as it depends on the deterministic network behaviors.

Stephen Dabideen and J.J. Garcia-Luna-Aceves [6], proposed that routing in MANET using depth first search (DFS) is feasible as well as efficient than breadth first search (BFS). The algorithm introduced is called as ordered walk search algorithm (OSA). The objective of this algorithm is to take advantage of the smaller time complexity of BFS and combine it with the low communication complexity of DFS in order to improve the efficiency of the search through the known path information. In order to demonstrate the effectiveness of OSA, ordered walk with learning (OWL) routing protocol has been presented which uses DFS to establish and repair paths from the source to the destination with minimum signaling overhead and fast convergence.

The following are the advantages of DFS over BFS that had been investigated by Stephen et al. in MANET:

(i) DFS require less overhead as compared to BFS. When large number of nodes is performing BFS, the routing overhead can saturate the network making it difficult to deliver any packets. However, DFS use only small network for routing.

(ii) When BFS is used in a network, where there are multiple flows of search packets this situation can lead to increase the load on network and loss of packets. On the other hand as DFS involves only a small part of the network, thus this reduces the load in the network and results in less packet loss.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Searching Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Breadth First Search</td>
</tr>
<tr>
<td></td>
<td>[6]</td>
</tr>
<tr>
<td>Overhead</td>
<td>More</td>
</tr>
<tr>
<td>Load in network</td>
<td>More</td>
</tr>
<tr>
<td>Packet loss</td>
<td>More</td>
</tr>
</tbody>
</table>

A comparative study of searching algorithms is shown in Table 1. From this table we conclude that DFS is better searching algorithm for MANET than BFS.

3. Proposed System

The block diagram of the proposed system is illustrated in Fig. 2. Each function of blocks is explained below:

1) Network Setup: Each node will send topology discovery packets to all its neighboring nodes. The aim of each node will be to setup connection with nodes available in the network.

2) Route Selection using Heuristic Approach: Heuristic approach require less time and space. The heuristic approach will be efficient as it will take feedback from data to find the path. Heuristic approach will start searching from source node to find traffic free path till it will reach destination node.

3) Statistical Traffic Analysis: The state of the network will be identified for certain period of time. This block will deduce point-to-point and end-to-end matrix.

4) Probability Distribution: The probability of a node being a source or destination is calculated in this block.
5) Source and Destination Probabilities: The result of a node as a source or destination will be interpreted graphically in this block.

One of the characteristic of MANET is that the identity of all the nodes is hidden. This proposed system will aim to unhide the identity of nodes by using statistical traffic analysis of the network. A heuristic search algorithm is used to search a particular node in the network. The searching algorithm chosen for searching the node will be depth first search (DFS). Source node will use DFS algorithm for traversing or searching the path in the network. Then statistical traffic analysis will be performed on these searched nodes. This analysis will provide an estimation of the data transmitted to all the neighboring nodes of every searched node. We can discover the traffic pattern by using probability distribution. The working of each of the module is explained in detailed below.

3.1 Searching Node in MANET Using Depth First Search

In this proposed system we are using DFS for routing decisions. DFS performs DFS search in distributed way in a given graph. DFS creates a path in the graph without creating jump from one node to another by restricting its neighbor node. When a node receives message for the first time, it sorts all its neighboring nodes according to their distance to destination and then uses that same order in DFS algorithm. It starts its searching from the source node and updates one hop neighbors. This search continues to reach traffic free path between source and destination node. The strategy of DFS is to explore deep in the graph.

In DFS all the child nodes are searched from the recent discovered source node. Those child nodes are searched whose source nodes were not earlier discovered. When all the child nodes have been searched, the search backtracks to discover the source node so that new source nodes can be explored. The process continues until all the child nodes discovered are reachable to source node. If still any child node remains undiscovered then any one of the node is selected as a new source node and this node search the undiscovered child nodes. The process continues until all nodes are found.

DFS is an efficient algorithm that can be used in routing decision for MANET. The following are some of the reasons:

(i) DFS performs with small number of nodes at a time. Thus routing in MANET can be carried out with less overhead. This will gradually increase the flow of number of packets.

(ii) As DFS involves only small number of nodes, there is no problem of flooding. Flooding of too many packets by large number of nodes can saturate the network. This leads to packet drop.

(iii) DFS reduces load in network and packet loss due to less flooding of packets.

These three reasons are strong enough to choose depth first search (DFS) as routing algorithm rather than breadth first search (BFS).

As shown in Fig. 3, depth first search works on tree or graph. The Fig. 3 gives an example of DFS routing path for the following graph: The searching starts from root node A. It is assumed that the left edges are selected than the right edges. Each node remembers the last visiting nodes which help to backtrack and reach the last node to complete the traversing. From Fig. 2 the path will be: A, B, D, H, I, E, C, F, J, G and K.

3.2 Statistical traffic analysis of packets in MANET

The traffic analysis of the network includes identifying the state of the network at each point of time. Time slicing technique is used to find out the state of network after regular intervals of time. Time slice technique is technique in which the process runs in a preemptive multitasking system. This time interval is called as the time slice or quantum. The scheduler runs every time slices for one time to choose the next process to run. The time slicing mechanism captures the snapshot of the network after each quantum time giving a n × n matrix where n is the number of nodes in the network. This traffic matrix will consists of traffic volume from one node to another.
For example,

\[
M = \begin{bmatrix}
0 & 1 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}
\]

Here 1 indicates that there is transmission of data (traffic volume) from node 1 to node 2 whereas 0 indicate that there is no transmission of data between the two nodes. Such matrices are generated after each time quantum \(M_1, M_2, M_3, \ldots, M_n\) which is used to find out the point to point probabilities of the nodes.

The traffic matrix obtained from the above helps to deduce point-to-point and end-to-end traffic volume between each pair of nodes. The adjacent matrices are added together to form a new matrix which is then added to the next adjacent matrix. Ex. \(M_1\) is added with \(M_2\) to form \(M_{1+2}\). \(M_{1+2}\) are then added with \(M_3\) to form \(M_{1+2+3}\) and so on. The final added matrix i.e. \(M_{1+n}\) is added with \(M_n\) to obtain final matrix which is termed as End to End Traffic matrix. Such a matrix will give one to one hope packets captured in all the matrices but will also device the two or more hope packet traffic between all the nodes in the network.

3.3 Discovery of System

The aim of the system is to discover the actual source/destination probabilities i.e. who are the actual sources and destinations of the communication taking place in the network. The source/destination probability calculation involves use of the final matrix \(M_n\). The matrix \(M_n\) is \(N \times N\) matrix where each row belongs to one node i.e. row 1 is expressing the traffic from node 1 to every other node in the network and so on. For computation of source/probability distribution following summation will be used which will give the probability of the selected node being a source/destination.

It is required to consider the identical probability distribution matrices should be initialized to \(S_0=D_0=(1/N, 1/N, 1/N, \ldots)\). The reason for employing the equal probability is that with no traffic taking place in the network all the nodes will have same chances of being a source and being a destination.

The equation required for source probability distribution is

\[
s'(i) = \sum_{j=1}^{N} r(i,j) \times d_0(j)
\]

The equation required for destination probability distribution is

\[
d'(i) = \sum_{j=1}^{N} r(j,i) \times s'(j)
\]

Where \(s(i)\) is probability of node \(i\) being source, \(N\) is number of nodes in the network, \(r(i,j)\) is the accumulative traffic volume from node \(i\) to node \(j\), \(d_0(i)\) is the destination vector.

In the Fig. 4, the flow of the proposed system is explained. When the system will start it will form a network. This network will consist of certain number of nodes. All the nodes will be browsed. In order to search the node a heuristic searching algorithm will be applied. If the required node is present then statistical traffic analysis will be performed on it. After performing statistical traffic analysis probability distribution will be applied to discover the traffic pattern. However, if the required node is not found then the system will stop and no further process will be carried out.

Fig. 4 System Flow Diagram
4. Proposed Solution

From the idea of the proposed system we are clear with two outcomes. These two outcomes are:

1) Discovery of traffic pattern in MANET
2) Probability of Source/Destination

5. Conclusion

The proposed system will observe the traffic pattern of the adversary. As nodes are hidden in MANET a heuristic searching algorithm will be applied. This heuristic searching algorithm will be depth first search (DFS). This system will perform statistical traffic analysis to find the data transmission between one to one and one to many nodes. Probability of point to point transmission among receivers will be estimated by point-to-point traffic matrix. Then by calculating multihop traffic and performing probability distribution the traffic pattern will be discovered. This will provide an approximate traffic pattern with approximate source and destination in the network. The proposed system will reduce the issue of anonymous communication in mobile ad hoc network (MANET).

References


Authors

Priyanka Sen received the BE degree in Computer Science and Engineering from Rashtrasant Tukadoji Maharaj Nagpur University in Maharashtra, India in 2013. She is currently working toward the ME degree in Wireless Communication and Computing from Rashtrasant Tukadoji Maharaj Nagpur University in Maharashtra, India. Her research interest includes ad hoc network and mobile ad hoc network security.