TPDS: A Heuristic Approach for Traffic Pattern Discovery System in Mobile Ad-hoc Network

Priyanka Sen¹, Vaishali Sahare²

^{1,2}Department of Computer Science and Engineering G.H Raisoni Institute of Engineering and Technology for Women, Nagpur

Abstract : There is absence of centralized coordination among nodes in Mobile Ad-hoc Network (MANET). This makes MANET vulnerable to attacks especially passive attacks. Security is one of the major concerns in military application. Anonymity hides the role of a node as source or destination. In order to trace the traffic pattern of adversaries this system is proposed. The objective of proposed system is to identify an approximate source and destination nodes present in the network. In MANET, identification of these nodes helps to trace the traffic pattern that is followed by an adversary. The pattern is traced by using best-first search (a heuristic approach) and discovered by statistical traffic analysis. Probability distribution is used to find the source and destination nodes. The implementation and analysis is carried out in Network Simulator 2 (NS2). Since the purpose of the proposed system is to trace the path of an adversary it is named as Traffic Pattern Discovery System (TPDS).

Keywords: Anonymity, mobile ad-hoc network (MANET), Network Simulator 2 (NS2), passive attack, statistical traffic analysis

I. Introduction

In mobile ad-hoc network (MANET), nodes are mobile in nature. This makes the topology dynamic. The wireless medium used has limited resource and also lack of centralized administration. Due to above characteristics, anonymous communication becomes a challenge in MANET. Many anonymous routing protocols such as Anonymous On-Demand Routing (ANODR) [1] and On-demand Lightweight Anonymous Routing (OLAR) [2] have been proposed for anonymity of nodes in network. Still adversary can intercept the information through passive attack.

MANET can be attacked in two ways: Active and Passive. Passive attack is more harmful than active attack as the user is unaware of collecting information performed by an adversary without disturbing the operation. However, MANET is vulnerable to traffic analysis that is one of the types of passive attack. Traffic Analysis is the process of monitoring and analyzing the activities of traffic patterns during delivery of packets in MANET. Traffic analysis [3] can be useful to carry a military operation. The soldiers can easily keep watch on their adversary through traffic analysis. Nodes that are within the transmission range can communicate directly, while communication between more than two nodes can take place through intermediate nodes. As centralized administration is absent in MANET, each node maintain their own routing and resource management in distributed manner. This makes MANET vulnerable to traffic analysis.

Traffic analysis is further subdivided into predecessor attack [4] and disclosure attack [5]. Due to the following nature of MANET above attacks cannot be used: 1) Broadcasting nature: Transmission of messages to multiple receivers result into uncertainty. 2) Ad-hoc nature: Difficult to determine role of each node in the network. 3) Mobile nature: Mobility of communication peers are not considered in traffic analysis model. The better option other than traffic analysis is statistical traffic analysis. Statistical traffic analysis collects information from statistical characteristics of traffic pattern. In statistical traffic analysis they neither modify nor delete the packets but they gather the packets and perform statistical traffic analysis on them.

There is requirement of such a technology which can perform statistical traffic analysis in MANET. The objective of this proposed system is to find an approximate source and destination nodes in order to trace the traffic pattern without the knowledge of adversary.

The remaining paper is organized in the following manner: Section II reviewed the earlier system. Section III introduces the proposed system. Section IV presents the implementation of the proposed system. Section V discusses the result analysis. While, section VI presents the conclusion.

II. Related Work

Qin et al. [6], proposed STAR (Statistical Traffic Pattern Discovery System) to trace traffic pattern in MANET. STAR uses statistical traffic analysis to find source and destination nodes. This proposed system have not mentioned searching algorithm to find out traffic less path.

Parameswaran et al. [7], projected comparative study of traffic analysis in MANET. They discussed on anonymity, traffic analysis, analyzed traffic analysis attacks and interpreted on wired network and MANET.

Lu and Liu [8], investigated the prolog engine whose purpose is to search a knowledge base to satisfy the given query. They concluded that Best-First Search is a better option when compared with depth first search (DFS) for Prolog engine. The conclusion is proved by following advantages considered by Lu et.al:

(i) In DFS solutions that are far away takes more time to reach desired path as compared to nearer solutions. This increases the execution time.

(ii) Best-First Search does not get trapped into endless branch of a search tree as DFS.

(iii) Efficiency of best-first search is better than DFS as that can be concluded from the above discussed issues. Table I shows the comparison of Best-First Search with DFS.

1 4010 1				
Parameters	Searching Algorithms			
	Best-First Search	Depth First Search		
Search time	Requires less time	Requires more time		
Result	Assured	Not assured		
Capability	Higher	Lower		

Table I. Comparison of Searching Algorithm

Due to following advantages Best-First Search is an appropriate searching algorithm for MANET: 1) Even in large problem space best-first search requires less time to generate solution. 2) Other paths are checked if selected path becomes less promising. 3) Low cost path is found out. 4) Considered as one of the options for complex problems.

Kelly et al. [9], reviewed on anonymity. Anonymity means services that can be kept invisible from adversary. This helps to protect user's data. Anonymity has three properties: Unidentifiability, Unlinkability, and Unobservability. Sender anonymity (SA), receiver anonymity (RA), mutual anonymity (MA) and group anonymity (GA) are subdivision of unidentifiability. Our focus is on unidentifiability in order to reduce sender anonymity (SA) and receiver anonymity (RA). This assists us to identify the traffic pattern.

Liu et al. [10], proposed anomaly detection system to detect the distributed denial of service (DDoS) attack in MANET. Two important metrics to detect DDoS attack are traffic intensity and packet number. These metrics are used whenever traffic analysis is carried out to detect the DDoS attack. Similarly data transmission is required for traffic analysis in our proposed system.

Zhang and Zhang [11], introduced the control traffic for route discovery in MANET. 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 are included to carry out analysis on control traffic. On-demand routing protocols are used to execute this research. Control traffic is affected by mobility, node density and data traffic. Theoretically data traffic is one of the factors of control traffic. Therefore, in route discovery control traffic also plays a major role.

Liu et al. [12], constructed traffic inference algorithm (TIA). In TIA, an adversary is allowed to interpret the traffic pattern in MANET. According to this algorithm passive adversary can notice the difference between data frames, routing frames and MAC control frames. These differences helps an adversary to determine MAC control frames using point-to-point traffic, recognize the routing frames by tracing end-to-end traffic and finally by using the data frames actual traffic pattern can be found. This algorithm depends on deterministic network behaviors is its major drawback.

Tehrani and Shahnasser [13], reviewed on anonymity and significance of anonymous communication protocol. They analyzed best anonymous routing protocol and examined the advantages and disadvantages of current solution, shortcoming and challenges that are not contributed by these protocols.

Sen and Sahare [14], discussed the traffic pattern discovery in MANET. A heuristic approach is used to trace the hidden nodes. For route discovery they studied depth first search (DFS). However, Table I proves that best-first search is better option than DFS.

III. Proposed System

The proposed system discloses the identity of the nodes. Nodes are identified through statistical traffic analysis. Best-First Search is one of the heuristic search algorithms which are used to trace a path in the network. Statistical traffic analysis and Probability distribution are performed on path that has been traced out. Statistical traffic analysis deduces source-destination matrix and traffic matrix. Probability distribution calculates the possibility of node being sender and receiver. The block diagram of the proposed system is illustrated in Fig. 1. Each phase of this proposed system along with the blocks is explained below:



Fig. 1 Block diagram for discovery of traffic pattern

1) Network Formation: The aim of this block is to connect each node with its neighbor nodes. Every node send hello packet to neighboring nodes.

2) Route Selection using Best-First Search: Best-First Search is used for traversing the nodes. This heuristic approach is used to trace a traffic pattern. Nodes are selected on the basis of traffic free path.

3) Statistical Traffic Analysis: The state of the nodes can be identified at that time. This block deduces source-destination matrix and traffic matrix.

4) Probability Distribution: The possibility of nodes being sender and receiver is calculated from the traced out path.

5) Probabilities of Source and Destination: This block interprets graphical analysis of nodes acting as source or destination node.

In order to understand the working of the proposed system each phase is explained in brief:

A. Network Formation Phase

This phase connects each node to all adjoining nodes available in the network. Each node sends hello packets to its neighboring nodes. The purpose of hello packet is to inform the presence of node to its nearby nodes. The network formation takes place for 1-hop and 2-hop neighbor nodes.

B. Route Selection using Best-First Search

In order to trace the path of an adversary, time is considered to be vital factor. Uninformed search algorithms require large time and space. These characteristic makes uninformed search algorithm unsuitable to trace the path. In such situation, it is desirable to use heuristic search algorithm. Heuristic search algorithm is defined as an algorithm that uses heuristic functions for searching a distinct node [16]. Table II shows the reason to choose best-first algorithm than other heuristic search algorithms.

Best-First Search is a combination of depth first search (DFS) and breadth first search (BFS). This algorithm traverses a graph to explore one or more than one goal nodes. In best-first search, solutions can be discovered in less time interval.

Heuristic Search Algorithms	Reasons
Generate-and-Test	If problem space is large it takes long time
	to generate solution.
Hill climbing	Other moves are never considered if one of
	the moves is selected.
Problem Reduction	Cannot assured about the path with lowest
	cost.
Constraint Satisfaction	Solves only crypt-arithmetic problems.
Means-Ends Analysis	Complex problems cannot be solved.

 Table II.
 Reason to Choose Best-First Search

C. Statistical Traffic Analysis in Network

The function of this phase is to analyze the environment of the network at that point of time. The analysis is carried out on statistical parameter. Here the parameter is traffic volume i.e. transmission of data. Through analysis of traffic volume matrices are generated between two or more nodes. This matrix is a n x n matrix where n refers to number of nodes in the network. For example,

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

1 indicates that there has been transfer of data from node 1 to node 3. Whereas 0 indicates that no data have been transferred. This matrix deduces traffic volume between two or more number of nodes. The matrix generated is traffic matrix where number of packets transmitted between n numbers of nodes can be analyzed.

The next procedure of the proposed system is to find source and destination from the traced route. Through probability distribution the calculation of source and destination node is carried out. The formula required to calculate source and destination probability are discussed below. It is necessary to consider that initial probability distribution matrices should be initialized to $S_0=D_0=(1/N, 1/N, 1/N...)$, where N refers to number of nodes. The main idea for this initial probability distribution works on the assumption that all nodes have same chance of being source and destination with no traffic taking place in the network.

Source probability distribution is calculated through

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

Destination probability distribution is calculated through

$$d_{1}(i) = \sum_{j=1}^{N} r(j, i) \times s'(j), \qquad (2)$$

Where s (i) refers to probability of node i being source, N refers number of nodes in the network, r (i, j) indicates accumulative traffic volume from node i to node j, $d_1(i)$ indicates destination vector.

Here, Fig. 2 illustrates the flow of the proposed system. With the execution of simulation the first step is formation of topology. In formation of topology every node sends hello packets to its adjoining nodes. This step is used to setup connection with each node available in the network. Second step is to browse the nodes. The purpose to browse the node is to check whether the nodes are traffic free or not. When all nodes are browsed, best-first search algorithm is applied to trace a path. As path has been traced out, next step is to perform statistical traffic analysis. This step finds out data transmission among nodes. After performing analysis, traffic matrix is produced. With generation of traffic matrix, the role of each node is estimated. Through probability distribution, the probability of each node as source and destination is calculated from the traced path. After completion of calculation the simulation stops. If the simulation cannot find out optimal route then the simulation ceases.



Fig. 2 Flow diagram of proposed system

D. Routing of Data

In this stage, nodes with high probability of source and destination are displayed from the traced path. As mathematical analysis is already deduced in previous stage, this stage focuses on graphical analysis. The parameters considered for graphical analysis are success rate for traffic pattern on different total number of nodes, success rate for validity of nodes being source or destination and probability distribution.

IV. Implementation

A. Environment for Simulation

The simulation is carried on popular simulation tool called as Network Simulator (Version 2) i.e. NS2. This simulation tool is based on event-driven. NS2 can be used to study the dynamic nature of communication network. This tool simulates protocols and network function in wired and wireless. C++ and Object-oriented

Tool command Language (OTcl) are the two languages of NS2. The function of C++ is to define the internal mechanism and act as backend. OTcl is used to set up simulation by assembling and configuring the scheduling discrete events as well as objects. OTcl acts as a front end [19]. The simulation parameters required for this proposed system is shown in Table III

Table III Simulation Parameters		
Parameter	Value	
Routing Protocol	DSR	
Simulation Time	20 seconds	
Simulation Area	800x800 m ²	
Number of Nodes	25, 45, and 65	
Traffic Type	UDP	
Pause Time	0.5 seconds	
Mobility	10 meter/sec	
Packet Size	512 bits	
Data Rate	512 kbps	
Mobility Model	Random waypoint	
MAC	802.11(a)	
Channel Type	Wireless channel	
Mobility Model	Random waypoint	
Queue Model	CMU queue	

Tahle	TIT	Simulation	Parameters
Lanc		Simulation	1 an announs

B. Outcomes

The simulation is performed on three networks consists of three total number of nodes. The total numbers of nodes are 25, 45 and 65. Here, only the output of 45 total numbers of nodes is displayed. The outcomes are displayed according to their phases discussed in Section III.

i) Network Formation Phase

The formation of network for 1-hop neighbors is displayed in Fig. 3. Each node in the network, act as a sender node which sends hello packets to its neighbor node. The black colour node indicates that the node is in state to start network formation for 1-hop neighbors. Every node sends omni-directional signals to all its adjoining nodes. The pink colour node indicates that the node had complete network formation for 1-hop distance.



Fig. 3 Formation of network for 1-hop neighbors.

Here, Fig. 4 displays formation of network for 2-hop neighbors. Node 30 performs network formation can be seen in the display. It acts as a sender and sends hello packets. Node 30 is in pink colour which indicates that the node is in state to start sending hello packets. Omni-directional antennas sends signal. The violet colour node indicates that the node had completed network formation phase.



Fig.4 Formation of network for 2-hop neighbors

ii) Route Selection using Best-First Search

The transfer of data between nodes is illustrated in Fig. 5 best-first search traces a traffic free path for data transmission. The traffic free path is detected on the basis of transmission of traffic volume (data packets) between the nodes. The traced path turns into red colour. The outcome displays the selected traffic path consisting of node number 12, 5, 11 and 23. The packets are transferred between these nodes.



Fig. 5 Transmission of data between nodes

Course Destination Matrix

iii) Statistical Traffic Analysis in Network

	Source-Destination Matrix					
(D	0	0	0		
4	4	2	1	4		
(0	0	0	0		
(0	0	0	0		
1	12	5	11	23	Node	Number

Fig. 6 Matrix of source-destination

Here in Fig. 6, displays the outcome of traffic analysis performed on the traced path. This matrix consists of probability distribution values calculated on traced path. The last row displays each node number of the detected path. Above every node the values are their respective probability distribution. These values are calculated by using equation (1) and equation (2). These values help to analyze the highest probability of node being source and destination.



The matrix of traffic is illustrated in Fig. 7. These values show the transmission of data (traffic volume) between two nodes. This matrix helps to analyze the number of data packets transmitted between two nodes. These values are calculated by performing summation of multiple point-to-point matrixes. The last row displays the node number of the detected path. Each node number is corresponds to its respective traffic volume. iv) Routing of Data



Fig. 8 Discovery of source and destination nodes.

The snapshot of last phase of the proposed system is displayed in Fig. 8. The outcome displays the source and destination node of the detected traffic pattern. These nodes are marked with red colour and labeled with green colour. The node number 12 is marked as source node, since it is having the highest probability of being sender. Similarly node number 23 is marked as destination node, as it is having the highest probability of being receiver.

V. Result Analysis

In order to perform analysis of the proposed system graphical analysis is performed. This gives an assurance of the correct implementation of the proposed ideas. The following are the outcomes of the graphical analysis.



The analysis of source node probability is displayed in Fig. 9. The simulation is performed on three different total numbers of nodes. They are 25, 45 and 65. This graph shows the overall probability of being source node for each node in the network. The X-axis represents "Nodes", while Y-axis represents "Probability". The red colour line is used for 25 nodes, green for 45 nodes and blue for 65 nodes. For 25 nodes, the highest probability is present between node numbers 10 to 15. For 45 nodes, the highest probability is present between node numbers 10 to 30. While, for 65 nodes the highest probability is present between node numbers 20 to 40.



Fig.10 Destination node probability analysis

The analysis of destination node probability is displayed in Fig. 10. The simulation is performed on three different total numbers of nodes. They are 25, 45 and 65. This graph shows the overall probability of being destination node for each node in the network. The X-axis represents "Nodes", while Y-axis represents "Probability". The red colour line is used for 25 nodes, green for 45 nodes and blue for 65 nodes. For 25 nodes, the highest probability is present between node numbers 10 to 20. For 45 nodes, the highest probability is present between node numbers 25 to 35. While, for 65 nodes the highest probability is present between node numbers 35 to 45.

The success rate for probability distribution is illustrated in Fig. 11. Success rate is calculated to find out the correctness of working of the proposed system. Success rate can be defined as percentage of success among a number of attempts. Here, graphical analysis is displayed for different number of nodes. The graphical analysis shows that how successfully simulation is carried out for three different numbers of nodes i.e. 25, 45 and 65. The X-axis represents "Different total number of nodes". While Y-axis represents "Success rate". The red line is used for 25 nodes, green for 45 nodes and blue for 65 nodes. From the graph it can be concluded that 45 nodes have the most successful success rate followed by 25 nodes and 65 nodes. The overall success rate obtained from the graph is 93.67%. In the graph some lines are in downfall position. This downfall can be seen in the simulation carried with total number of nodes as 65.

The reason for this downfall is due to increase in number of nodes. As the number of nodes in the network increases the system has to analyze more number of nodes. This result in increment in the number of packets transferred in the network. With increment in the number of nodes, there is increase in performing analysis. This definitely increase the time required to perform the traffic analysis. The success rate is inversely proportional to number of nodes. With increase in number of nodes the success rate falls down. The formula for success rate used in this proposed system is:

Success rate = $\frac{Success}{No. of attempts}$

Where "Success" refers to the success of getting same node number multiple times as source or destination and "No. of attempts" refers to the number of times the simulation has been executed.



Fig.11 Success rate for probability distribution.

The success rate for different traffic pattern is illustrated in Fig. 12. Success rate is calculated to find out the correctness of working of the proposed system. Here, graphical analysis is displayed for different traffic pattern. The graphical analysis shows that how successfully simulation is carried out for three different numbers of nodes i.e. 25, 45 and 65. The X-axis represents "Different total number of nodes". While Y-axis represents "Success rate". The red line is used for 25 nodes, green for 45 nodes and blue for 65 nodes. From the graph it can be concluded that 45 nodes and 65 nodes have the most successful success rate followed by 25 nodes. The overall success rate obtained from the graph is 78.33%. This graphical analysis is performed to find out which nodes performs more successfully with different traffic pattern.



Fig.12 Success rate for different traffic pattern

VI. Conclusion

The main purpose of the proposed system is to disclose the identity of source and destination nodes in the network. This can be fulfilled by discovery of communication pattern. The communication pattern is discovered without decrypting the packets. This has been satisfied by using best-first search (a heuristic approach) for traversing the path, statistical traffic analysis for analyzing and to identify point-to-point transmission among receivers. This is followed by calculating probability distribution to find out approximate source and destination nodes in the traced path. This reduces anonymous communication which is one of the characteristics in mobile ad hoc network (MANET). The overall success rate obtained from the graph of success rate for probability distribution is 93.67%, while overall success rate obtained from the graph of success rate for different traffic pattern is 78.33%.

Best-First Search is one of the heuristic search algorithms that have been utilized in this proposed system. However, this best-first search has a drawback. Best-First Search algorithm terminates when no optimal path is found. This drawback can create an obstacle to find out the communication path of the adversary. Thus, in future instead of best-first search A* or AO* can be used as a heuristic search approach.

References

- J. Kong, X. Hong, and M. Gerla, "An Identity-Free and On- Demand Routing Scheme against Anonymity Threats in Mobile Ad Hoc Networks," IEEE Trans. Mobile Computing, vol. 6, no. 8, pp.888-902, Aug. 2007.
- [2] Y. Qin and D. Huang, "OLAR: On-Demand Lightweight Anonymous Routing in MANETs," Proc. Fourth Int'l Conf. Mobile Computing and Ubiquitous Networking, pp. 72-79, 2008.
- [3] J. Raymond, "Traffic Analysis: Protocols, Attacks, Design Issues, and Open Problems," Proc. Int'l Workshop Designing Privacy Enhancing Technologies: Design Issues in Anonymity Unobservability, pp. 10-29, 2001.
- [4] M. Wright, M. Adler, B. Levine, and C. Shields, "The Predecessor Attack: An Analysis of a Threat to Anonymous Communications Systems," ACM Trans. Information and System Security, vol. 7, no. 4, pp. 489-522, 2004.
- [5] G. Danezis, "Statistical Disclosure Attacks: Traffic Confirmation in Open Environments," Proc. Security and Privacy in the Age of Uncertainty, vol. 122, pp. 421-426, 2003.
- [6] Yang Qin, Dijiang Huang and Bing Li, "STARS: A Statistical Traffic Pattern Discovery System for MANETs" IEEE Transactions on Dependable and Secure Computing, Vol. 11, No. 2, March/April 2014.
- [7] T. Parameswaran, Dr. C. Palanisamy, M.Karthigadevi, "Comprehensive Study of Traffic Analysis In MANET", International Journal of Research in Advent Technology, Vol.2, No.10, pp. 151-159, October 2014
- [8] Benjie Lu and Zhingqing Liu,"Prolog with Best First Search", IEEE 25th Chinese Control and Decision Conference, 2013.
- [9] Douglas Kelly, Richard Raines, Rusty Baldwin, Michael Grimaila, and Barry Mullins, "Exploring Extant and Emerging Issues in Anonymous Networks: A Taxonomy and Survey of Protocols and Metrics", IEEE Communications Surveys & Tutorials, Vol. 14, No. 2, Second Quarter 2012.
- [10] Lei Liu, Xiaolong Jin, Geyong Min, and Li Xu, "Real-Time Diagnosis of Network Anomaly based on Statistical Traffic Analysis", IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, 2012.
- [11] Zhilin Zhang and Yu Zhang, "Control Traffic Analysis of On-Demand Routing Protocol in Mobile Ad-hoc Networks", IEEE Second International Conference on Networking and Distributed Computing, 2011
- [12] Y. Liu, R. Zhang, J. Shi, and Y. Zhang, "Traffic Inference in Anonymous MANETs," Proc. IEEE Seventh Ann. Comm. Soc. Conf. Sensor Mesh and Ad Hoc Comm. and Networks, pp. 1-9, 2010.
- [13] Tehrani, A.H. and Shahnasser, H.," Anonymous communication in MANET's, solutions and challenges," IEEE International Conference on Wireless Information Technology and Systems (ICWITS), pp. 1-4, 2010.
- [14] Priyanka Sen and Vaishali Sahare, "A Survey on Traffic Pattern Discovery in Mobile Ad hoc Network," International Journal of Computer Science and Network (IJCSN) Vol.4, No. 1, pp. 25-30, 2015.
- [15] Src (Fig.1): Block Diagram for Discovery of Traffic Pattern: Priyanka Sen and Vaishali Sahare, "A Survey on Traffic Pattern Discovery in Mobile Ad hoc Network," International Journal of Computer Science and Network (IJCSN) Vol.4, No. 1, pp. 25-30, 2015.
- [16] Elaine Rich, Kevin Knight and Shivashankar B Nair, Artificial Intelligence, Third Edition, Tata McGraw Hill, 2009.
- [17] Src (for equation (1)): Yang Qin, Dijiang Huang and Bing Li, "STARS: A Statistical Traffic Pattern Discovery System for MANETs" IEEE Transactions on Dependable and Secure Computing, Vol. 11, No. 2, March/April 2014.
- [18] Src (for equation (2)): Yang Qin, Dijiang Huang and Bing Li, "STARS: A Statistical Traffic Pattern Discovery System for MANETs" IEEE Transactions on Dependable and Secure Computing, Vol. 11, No. 2, March/April 2014.
- [19] Teerawat Issariyakul and Ekram Hossain, Introduction to Network Simulator NS2, Second Edition, Springer, 2012.