International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 10 (2015) pp. 26605-26616

http://www.ripublication.com

A Generalized Statistical Traffic Pattern Discovery System for MANETs

P.Mounika

M.Tech II year, Dept of CSE

SreeVidyanikethan Engineering college

Tirupathi, Andhra Pradesh

mounikapaduchuru@gmail.com

Mr.S.Bavaji

M.Tech, Assistant professor, Dept of CSE

SreeVidyanikethan Engineering college

Tirupathi, Andhra Pradesh

bavaji.shaik2000@gmail.com

Abstract:

This paper works passively attacks the network to perform traffic analysis by identifying source and destination probability distribution. In STARS, it is difficult to analyze the traffic over geographical area. The sensors cannot analyze the traffic when nodes are close to each other. For better traffic analysis, generalized statistical traffic pattern discovery system for MANETs is proposed. In GSTARS, the network is divided into clusters and traffic analysis is done in each cluster using STARS. This method gives better traffic analysis in terms of probability distribution than STARS.

Index terms: Anonymous communication, mobile ad hoc networks, statistical traffic analysis, point-to-point traffic matrix, end-to-end traffic matrix.

1 INTRODUCTION

MANETS is a multi-hop routing network where one node tries to send information to other nodes which is out of its communication range. The packets are forwarded via intermediate nodes without the use of any centralized authority. The attacks of MANET is classified into active attacks and passive attacks. Passive attack monitors the network traffic only and does not alter the data i.e., transmitted within the network. It does not interrupt the operation of a routing protocol. To detect this type of attack is very difficult because the operation of network doesn't get affected. Hence to overcome this, powerful encryptions algorithms are used to encrypt the data.

Mobile ad hoc networks are severely important in real time and military communications. Hence, it requires anonymous communications to prevent wireless passive attacks. Anonymous routing protocols are used such as ANODR, ALARM, MASK, onion routing to achieve anonymous communications of MANETS.

Communication anonymity is a major issue in MANETs, which consists of the following:

i. **Source/Destination anonymity**: It is difficult to find the sender and receiver movements of the network.
ii. **End-to-End relationship anonymity**: It is difficult to find the end-to-end communication relations.

Huang proposed Evidence based statistical traffic analysis model for MANETs. Snapshot is triggered when one node sends a packet to another node and construct point-to-point traffic matrix for each hop. The end-to-end traffic matrix is originated through series of point-point traffic matrix. The first scheme fails to address the maximum hop-count of a packet when deriving end-to-end traffic matrix. It doesn't provide any method to define the source/destination probability distribution.

By re-cycling, the evidence-based statistical model, STARS is proposed. The main of STARS is to derive the source/destination probability distribution and end-to-

end communication relations.

2 RELATED WORK

Traffic analysis is applied in military and real time applications.Hence,this models have been used against the static wired networks. In brute force attack, the dummy messages are sent through the network that gets complex the adversary process. The brute force attack can be carried out by passive, external attack which are to be applied in wired networks only.

The predecessor attack doesn't effectively analyze the traffic because of the following constraints:

i. **The broadcast nature**: It is difficult to detect the destination accurately because the packets sent and received by many nodes.

ii. **The ad hoc nature**: The ad hoc networks are infrastructure less network. Hence, each node acts as both sender and receiver.

iii. **The mobile nature**: In MANETs, the nodes are free to move with different speeds. The network topology is dynamically changing due to the mobile nature of nodes. Hence, difficult to analyze the communication between nodes.

Evidence based statistical Traffic analysis build a point-to-point transmission hops based on raw traffic. Use these matrices to develop the end-to-end traffic matrix. This method doesn't suitable for multi-hop traffic. Lot of inaccuracies is present.

3 EXISTING WORK

STATISTICAL TRAFFIC PATTERN DISCOVERY SYSTEM

STARS involve mainly two steps

i. Construct the point-to-point traffic matrices and then build end-to-end traffic matrix.

ii. Based on end-to-end traffic matrix, calculate source and destination probability for each node and relationship between the source and destination.

Consider the following scenario Fig:1

In Fig:1 Node 1 which is not in the range of Node 3.Hence it broadcast the packet

through the intermediate node i.e., Node 2. Node 2 is in transmission range of Node 1 and Node 3. Therefore, it acts as a neighborhood node for both nodes.

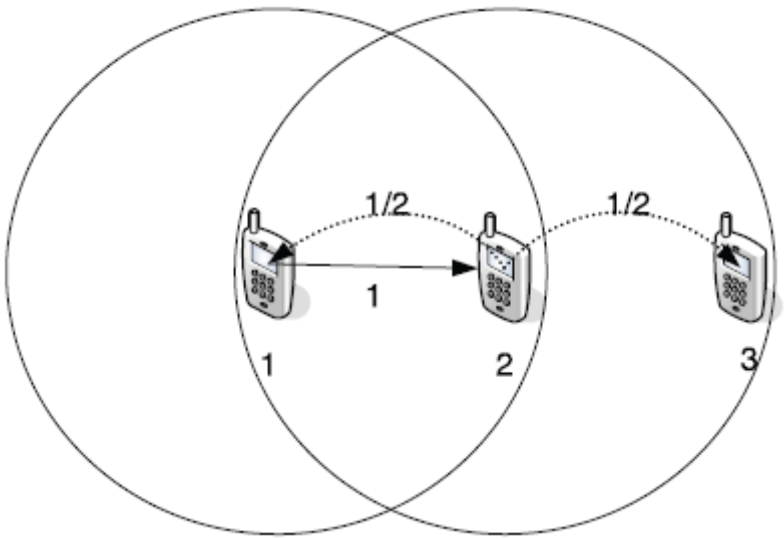

Fig: 1. A simple ad hoc network

3.1Traffic Matrices Construction

3.1.1 Point-to-point traffic matrix:

First apply a time-sharing technique to the above Fig:1 that is a snapshot is produced while one node sends a packet to another node. we need to create the point-to-point traffic matrices for each hop. The matrices construction for each hop such as$M_{1\times e} = (M_1, M_2, \ldots\ldots M_k)$. Traffic matrix $M_e = (m_e(a,b))_{N\times N}$ is created for direct communication between the nodes with in the interval. If there is a direct communication link from node a to node b then, the $m_e(a,b)$= 1. If the node b which moves out of the communication range of node a then $m_e(a,b)$=0. The virtual size of a packet is "1".For a one node, if there is a lot of neighborhood nodes are present then the virtual size for each sub –packets size is "1/n". The traffic matrices is constructed for each hop is given below

$$M_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} M_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0 & 0 \end{bmatrix}$$

Here the node2 which acts as a neighborhood node for node 1 and node 3 since it broadcast a real packet from node 1 to node 3. The node2 forwards a packet to node1 and node 3 that is split it into two subpackets. Hence, its virtual size for sub packets is 0.5. M is mass matrix of each node.

3.1.2 End-to-End Traffic Matrix:

Calculate the end-to-end traffic matrix E=$(e(a,b))_{N\times N}$ based on sequence of point-to-point traffic matrix, where $e(a,b)$ is acquisitive traffic volume from node a to node b.

Consider the following algorithms for point-to-point and end-to-end traffic matrix.

Algorithm 1- point –to-point traffic matrix *f*($M_{1\times k}$)

1: E=M_1

2: for e=1 to K-1 do

3: E= g(E, M_{e+1})+M_{e+1}

4: end for

5: return E

Algorithm 2- End-to-End traffic matrix g(E,M_{e+1})

1:$E' = E$

2: for a=1 to N do

3: for k=1 to N and k ≠ i do

4: for b=1 to N do

5: for each lє M_{e+1}(b,k).pkt do

6: if □ mє r(a, b).pkt s.t .*l.time-m.time*< T andm.hop<H then

7: create i with i.time = l.time

i.hop= m.hop+1

i.vsize= min{l.vsize, m.vsize}

8: $e'(i,k).pkt = r'(i,k).pkt \cup \{z\}$

9:$e'(i,k) = r'(i,k) + i.vsize$

10: end if

11: end for

12: end for

13: end for

14: end for

15: return E'

Where E is an end-to-end traffic matrix that is initiated through point-to-point traffic matricesM_1 to M_e. M_{e+1} iterative of point-to-point traffic matrix.

One constraint is the traffic volume from node a to node b should not exceed the volume of any transmissions. Hence, we have $Pd_{1,3}.vsize = min\{Pd_{1,2}.vsize, Pd_{2,3}.vsize\}$=0.5. Finally we get, the end-to-end traffic matrix is as follows

$$E=\begin{bmatrix} 0 & 1 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0 & 0 & 0 \end{bmatrix}$$

Where E is an end-to-end traffic matrix.

3.2 Traffic Pattern Discovery

To discover the actual sender and receiver, still need to calculate source/destination probability distribution and end-to-end link probability distribution.

3.2.1 Source/Destination probability Distribution

We denote the actual sender/source and receiver/destination probability distribution as two vectors S and R respectively. To compute S and R, the sender probability distribution vector series $\overline{S}$=(,S_0,S_1, ….S_n,….), and receiver vector probability distribution is $\overline{R} = (R_0, R_1, \ldots, R_n, \ldots)$

Deprived of any traffic information default all nodes are likely to be sender

and receiver. The uniform probability distribution vectors is S_0=R_0=(1/N,1/N,….,1/N).

Consider the equation (1) , In matrix E, the ath row (r(a,1)…r(i,N)) from node 'a' to every node in wireless MANET. If we multiply this vector by R_0, we get

$$s'(a) = \sum_{b=1}^{N} e(a,b) \times r_0(b) \quad (1)$$

Where b=1,2,3…n

The node itself has high probability of being a sender when a node transmits a lot of packets to another node with being an receiver.

Similarly for receiver, it uses the result from (1) and multiplying the ath row in the transpose of E by S', we get

$$r_1(i) = \sum_{b=1}^{N} e(b,a) \times s'(b) \quad (2)$$

Where b=1,2,3..n

The node itself has high probability of being a receiver when a node receives a lot of packets from another node with high probability of being a sender.

According to the above analysis, the iterative algorithm for sender and receiver is

$$S_{n+1} = (E.E^T).S_n \quad (3)$$

and $$R_{n+1} = (E.E^T).R_n \quad (4)$$

Based on this formulas, we get actual sender and Receiver values. For node which is having the highest probability of source that is the actual source. Similarly, for which is having the highest probability of receiver that is the actual destination.

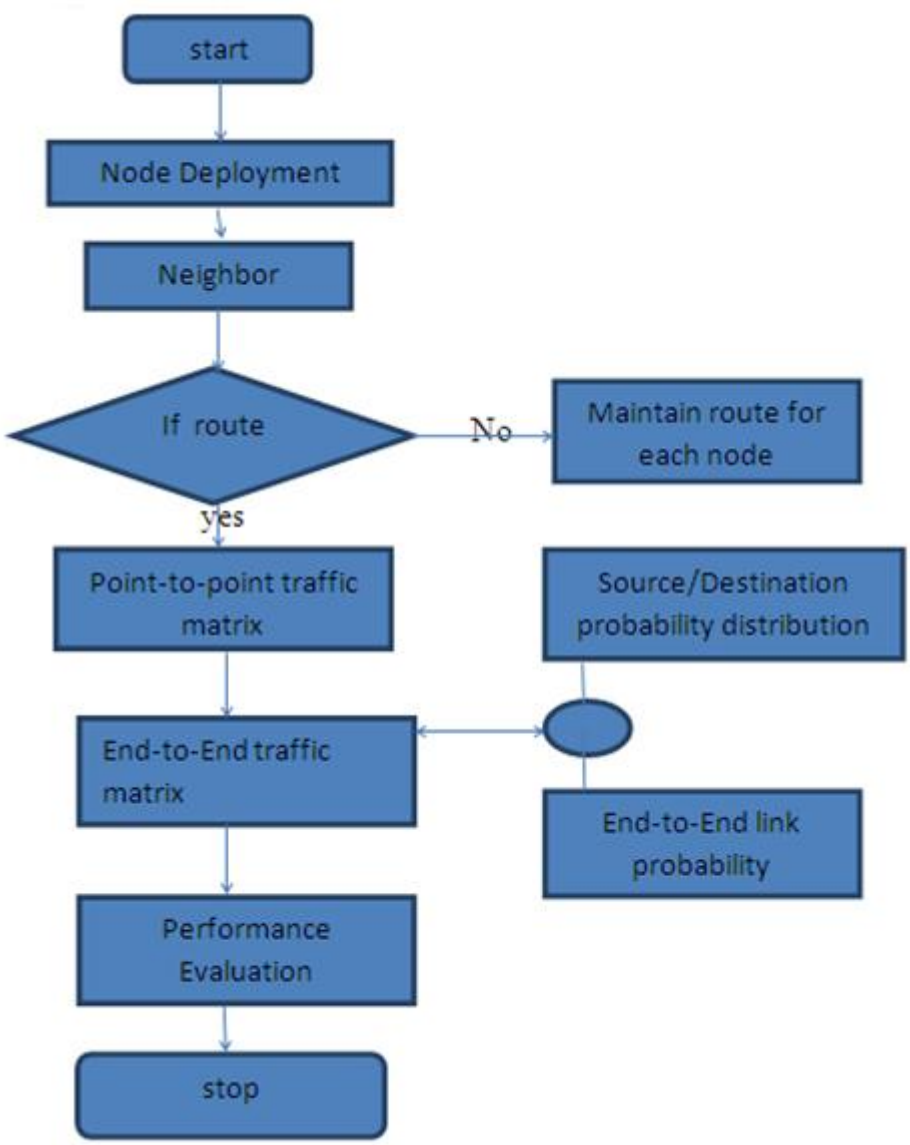

Fig: 2.Work flow of analyzing the traffic

3.2.2 End-to-End link probability distribution:

If the traffic flow is from node a to node b and node a is likelihood to be a sender , ignore the traffic which comes out from node a ,then we set the value is zero. By eliminating the traffic which comes from node a, some of the nodes which links to the node a is also dropped that means the end-to-end link is present is between the node a to node b.

4 PROBLEM STATEMENT

It is difficult to analyze the traffic over global area for the passive assailant i.e., to

identify the actual source and destination.

5 IMPLEMENTATION

For effectively performing the traffic analysis, the network is divided into regions and traffic analysis is done in each region using STARS. Hence, Generalized statistical traffic pattern discovery system is used for MANETs, it figure out the actual source and destination and relationship between them.

- Create a MANET comprised of 30 mobile nodes deployed in a $800 \times 800m^2$
- Divide the MANET into regions by using clustering.
- Construct the point-to-point and end-to-end traffic matrices.
- Analyze the traffic by using point-to-point traffic matrices and end-to-end traffic matrix.
- Identify the actual source and destination.

5.1 Cluster Formation

Cluster is formed based on CGSR algorithm as shown in Figure3. The nodes which are geographically nearby is aggregated in the cluster. The node itself chosen as CH and it sends a request message to all nodes in the cluster. The cluster head gateway switch (CGSR) uses a distributed algorithm called the Least cluster change (LCC).The LCC algorithm is considered stable since the cluster head will change only under two conditions. When two cluster heads come within the range of each other or when a node gets disconnected from any other cluster. In cluster, first source Communicate to the cluster head and its communicate to the another cluster head via intermediate nodes of two clusters and then finally CH to the destination node.

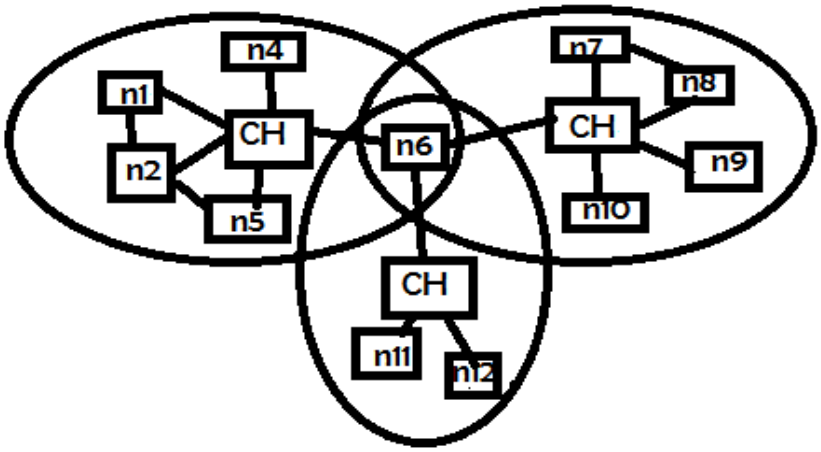

Fig. 3.Clustering

Simulation Parameters

No. of nodes	25
Area	1000 *1000 m^2
Routing protocol	AODV
Traffic Source	CBR
Mobility model	Random way point
T(s)	1.0
Node speed	10m/s
Transmission rate	11 Mbps
Packet size	512 bytes

6 EXPERIMENT

The output of STARS is based on probability distributions. The MANET which consists of 30 mobile nodes employed in a 1000 *1000 m^2 area. The probability distribution is used to figure out the actual source and destination. The network environment is simulated in NS2.

7 RESULTS

The GSTARS which gives better traffic analysis, it figure out the actual source and destination nodes than STARS. The performance is increased in terms of Probability distribution. The analysis for source probability distribution and destination probability distribution as shown below

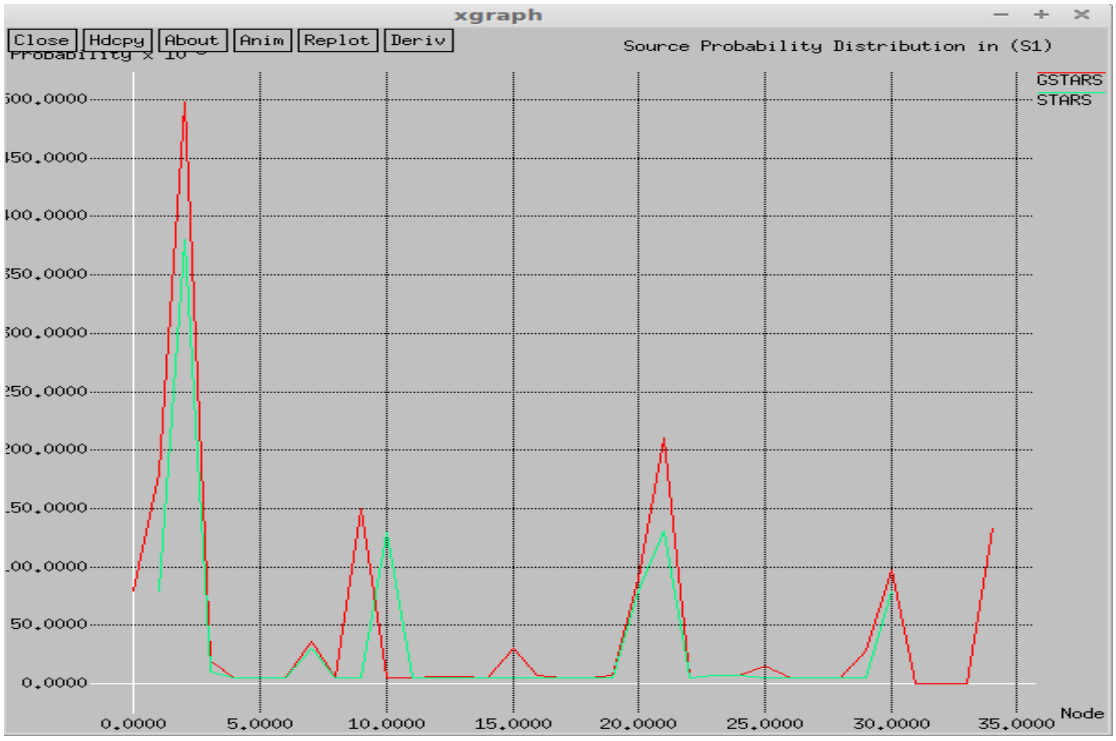

Fig. 4. Source probability distribution

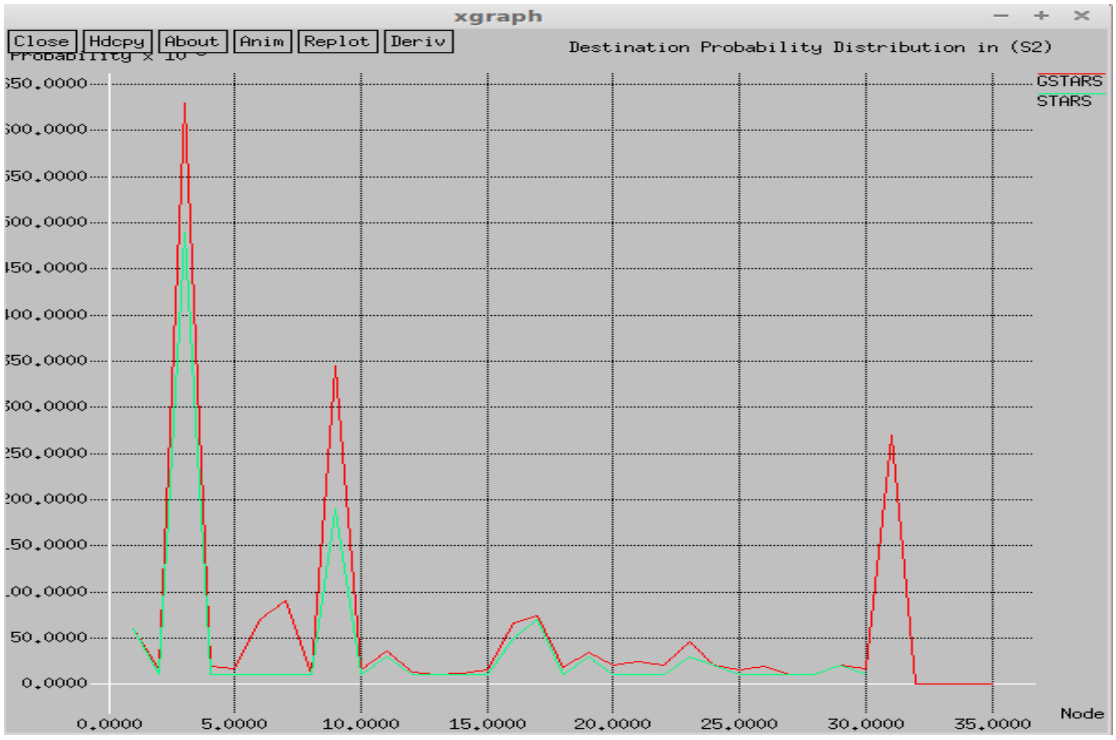

Fig. 5. Destination probability distribution

8 CONCLUSION

The passive attack is to determine the sources and destination using GSTARS, where the entire network is divided into clusters. Apply STARS technique inside the clusters. Based on the captured traffic it constructs a point-to-point traffic matrix and end-to-end traffic matrix. It uses end-to-end traffic matrix to derive probability distribution. The probability of becoming certain node as a source and certain node as a destination in GSTARS is better than STARS.

REFERENCES

[1] **Yang Qin, Dijiang Huang, *"STARS: A Statistical Traffic Pattern Discovery System for MANETs,"* IEEE Trans. On Dependable and Secure Computing., vol.11, no.2, March/April 2014**

[2] **M. Reed, P. Syverson, and D. Goldschlag, "*Anonymous Connections and Onion Routing*," IEEE J. Selected Areas in Comm., vol. 16, no. 4, pp. 482-494, May 2002.**

[3] **G. Danezis and A. Serjantov, "*Statistical Disclosure or Intersection Attacks on Anonymity Systems,"* Proc. Sixth Information Hiding Workshop (IH '04), pp. 293-308, 2004.**

[4] **T. He, H. Wong, and K. Lee, *"Traffic Analysis in Anonymous MANETs,"* Proc. Military Comm. Conf. (MILCOM '08), pp. 1-7, 2008.**