International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 5, May 2015)

# STATISTICAL TRAFFIC ANALYAIS MODEL FOR ANONYMOUS MANETs

Shashikanth Nayak
Computer Network Engineering
Dayananda Sagar College of Engineering Bangalore
shashiguru07@gmail.com

Dr. Amutha.S
Professor
Dayananda Sagar College of Engineering,
Bangalore

**Abstract - Anonymous Communication is the important issue in case of Mobile ad hoc Networks (MANETs). It is very difficult to find the source and destination of the communication link and the other intermediate nodes involved in it. Many enhancement techniques are proposed to enhance the anonymous communication in case of the mobile ad hoc networks (MANETs). However, MANETs are vulnerable under certain circumstances like passive attacks and traffic analysis attacks. Here we proposed the traffic analysis model, expose some of the methods and attacks that could infer MANETs are still weak under the passive attacks. To show how to discover the communication patterns without decrypting the captured packets, we present the paper Statistical traffic analysis model. In order to discover the packet patterns this model works passively and does the traffic analysis based on the statistical characteristics of the captured raw traffic. Here we can determine the source node, destination node and the end-to-end communication path in case of mobile ad hoc networks. And also accuracy of revealing the hidden traffic is more than other existing systems.**

***Keywords – MANETs, Anonymous Communication, Statistical Traffic Analysis, Time Slicing.***

## I. INTRODUCTION

A mobile ad hoc network (MANET) is a continuously self-configuring infrastructure less network of mobile devices connected without wires. In the mobile ad hoc network, nodes can directly communicate with all the other nodes within their radio ranges; whereas nodes that not in the direct communication range use intermediate node(s) to communicate with each other. In these two situations, all the nodes that have participated in the communication automatically form a wireless network, therefore this kind of wireless network can be viewed as mobile ad hoc network.

MOBILE ad hoc networks (MANETs) are originally designed for military tactic environments. Communication anonymity is a critical issue in MANETs, which generally consists of the following aspects: 1) Source/destination anonymity: it is difficult to identify the sources or the destinations of the network flows. 2) End-to-end relationship anonymity: it is difficult to identify the end to- end communication relations. To achieve anonymous MANET communications, many anonymous routing protocols such as ANODR [1], MASK [2], and OLAR [3] have been proposed. Though a variety of anonymity enhancing techniques like onion routing and mix-net are utilized, these protocols mostly rely on packet encryption to hide sensitive information (e.g., nodes' identities and routing information) from the adversaries. However, passive signal detectors can still eavesdrop on the wireless channels, intercept the transmissions, and then perform traffic analysis attacks. Over the past few decades, traffic analysis models have been widely investigated for static wired networks. For example, the simplest approach to track a message is to enumerate all possible links a message could traverse, namely, the brute force approach. Recently, statistical traffic analysis attacks have attracted broad interests due to their passive nature, i.e., attackers only need to collect information and perform analysis quietly without changing the network behavior (such as injecting or modifying packets). The predecessor attack and disclosure attacks are two representatives. However, all these previous approaches do not work well to analyses MANET traffic because of the following three natures of MANETs: 1) the broadcasting nature: In wired networks, a point-to-point message transmission usually has only one possible receiver. While in wireless networks, a message is broadcasted, this can have multiple possible receivers and so incurs additional uncertainty. 2) The ad hoc nature: MANETs lack network infrastructure, and each mobile node can serve as both a host and a router. Thus, it is difficult to determine the role of a mobile node to be a source, a destination, or just a relay. 3) The mobile nature: Most of existing traffic analysis models does not take into consideration the mobility of communication peers, which make the communication relations among mobile nodes more complex.

In past, enhancing techniques of MANETs [4] are based on packet encryption to protect the communication anonymity of the mobile ad hoc networks. In Evidence-based statistical traffic analysis approach provides a practical attacking framework against MANETs but still leaves substantial information about the communication patterns undiscovered. In this approach, the system for discovering a statistical traffic in MANETs is capable of discovering the sources, the destinations, and the end-to-end communication relations.

Finally using the performance evaluation technique comparing collected traffic pattern with actual traffic pattern to calculate the false positive rate and false negative rate based on typical confusion matrix and this procedure calculates the accuracy in disclosing the hidden traffic patterns.

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 5, May 2015)

## II. SYSTEM DESIGN

The system design details and modules of each design are shown in the fig.1 below. And this section also explains the algorithms for each module for statistical traffic analysis model.

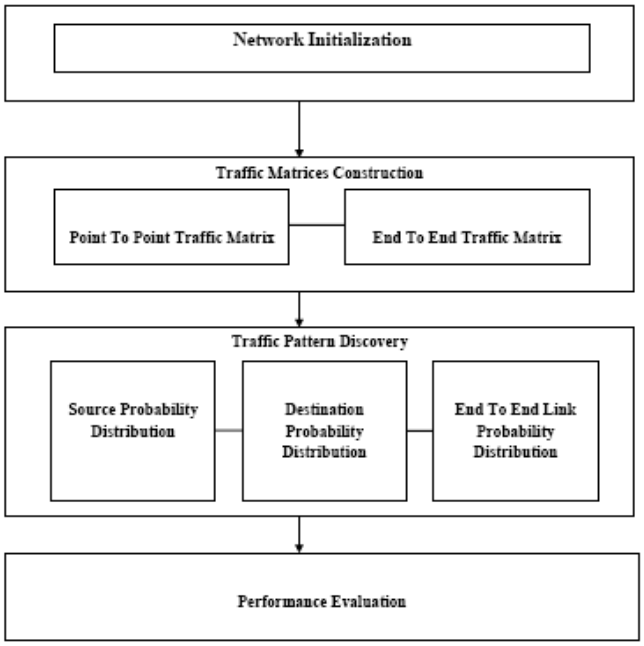


Fig.1 System Design

In Fig.1 shows there are four modules for statistical traffic analysis model for anonymous MANET. They are Network Initialization, Traffic Matrices Construction, Traffic pattern discovery system and finally performance evaluation.

### 1. Network Initialization

Network initialization is the first step in any network system design. In this module wireless network is initialized. In network initialization number of nodes, area of the network, nodes mobility model can be initialized. In network initialization module there are two models one is communication model and other one is adversary model. The assumptions for those models given as the PHY/MAC layer are controlled by the commonly used 802.11(a/b/g) protocol. But all MAC frames (packets) are encrypted so that the adversaries cannot decrypt them to look into the contents. Padding is applied so that all MAC frames (packets) have the same size. Nobody can trace a packet according to its unique size. The "virtual carrier sensing" option is disabled. The Source/destination addresses in MAC and IP headers are set to a broadcasting address (i.e., all "1") or to use identifier changing techniques. In this case, adversaries are prevented from identifying point-to point communication relations. No information about the traffic patterns is disclosed from the routing layer and above. Dummy traffic and dummy delay are not used due to the highly restricted resources in MANETs. The attackers' goal is to discover the traffic patterns among mobile nodes. Particularly, we have the following four assumptions for attackers: The adversaries are passive signal detectors, i.e. they are not actively involved in the communications. They can monitor every single packet transmitted through the network. The adversary nodes are connected through an additional channel which is different from the one used by the target MANET. Therefore, the communication between adversaries will not influence the MANET communication. The adversaries can locate the signal source according to certain properties (e.g., transmission power and direction) of the detected signal, by using wireless location tracking techniques [5] such as triangulation, nearest sensor, or RF fingerprinting. Note that none of these techniques can identify the source of a signal from several nodes very close to each other. Hence, this assumption actually indicates that the targeted networks are sparse in terms of the node density. In other words, any two nodes in such a network are distant from each other so that the location tracking techniques in use are able to uniquely identify the source of a wireless signal. In the following of this paper, unless specifically denoted as "signal source" or "source of signal," the word "source" indicates the source of a network flow. The adversaries can trace the movement of each mobile node, by using cameras or other types of sensors. In this case, the signals (packets) transmitted by a node can always be associated with it even when the node moves from one spot to another.

### 2. Traffic Matrices Construction

To disclose the hidden traffic patterns in a MANET communication system, this system includes two major steps. First, it uses the captured traffic to construct a sequence of point to point traffic matrix matrices and then derives the end to end traffic matrix. In traffic matrix construction there are two main steps to follow. One is finding the point to point traffic and other one is to deriving the end to end traffic matrix.

Capturing point to point traffic in certain period T. Using captured traffic; build point to point traffic matrices such that each traffic matrix only contains independent one hop packets. To avoid a single point to point matrix from containing two dependent packets, then apply time slicing technique[6]. Take a snapshot of the network during time interval $\Delta te$ and each snap shot is triggered by a captured packet. Constructs a traffic matrix which is an N*N one hop traffic relation matrix. A length of each time interval $\Delta te$ is determined by two criteria: one is, a node can be either sender or receiver within this time interval. But it cannot be both. And the second one is each traffic matrix must correctly represent the one-hop transmissions during the corresponding time interval and avoiding same packets from different nodes. Using point to point traffic matrix, find end to end matrix. The point to point traffic captured directly and multi hop traffic

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 5, May 2015)

deduced from the point-to-point traffic. We can implement the point to point traffic using algorithm1 using captured traffic in the network.

**Algorithm 1.** —$f(\mathbf{W}|_{1\times K})$.
1: $\mathbf{R} = \mathbf{W}_1$
2: **for** $e = 1$ to $K - 1$ **do**
3: $\quad \mathbf{R} = g(\mathbf{R}, \mathbf{W}_{e+1}) + \mathbf{W}_{e+1}$
4: **end for**
5: **return** R

Algorithm1 Finding Point to Point Matrix

After finding the point to point traffic we have to derive end to end traffic using the point to point traffic matrix technique. The algorithm for end to end traffic matrix is shown in Algorithm2 above.

**Algorithm 2.** —$g(\mathbf{R}, \mathbf{W}_{e+1})$.
1: $\mathbf{R}' = \mathbf{R}$
2: **for** $i = 1$ to $N$ **do**
3: $\quad$ **for** $k = 1$ to $N$ **and** $k \neq i$ **do**
4: $\quad\quad$ **for** $j = 1$ to $N$ **do**
5: $\quad\quad\quad$ **for each** $x \in w_{e+1}(j,k).pkt$ **do**
6: $\quad\quad\quad\quad$ **if** $\exists\, y \in r(i,j).pkt$ **s.t.** $x.time - y.time < \mathcal{T}$ **and** $y.hop < \mathcal{H}$ **then**
7: $\quad\quad\quad\quad\quad$ create $z$ with $z.time = x.time$, $z.hop = y.hop + 1$, $z.vsize = \min\{x.vsize, y.vsize\}$
8: $\quad\quad\quad\quad\quad$ $r'(i,k).pkt = r'(i,k).pkt \cup \{z\}$
9: $\quad\quad\quad\quad\quad$ $r'(i,k) = r'(i,k) + z.vsize$
10: $\quad\quad\quad\quad$ **end if**
11: $\quad\quad\quad$ **end for**
12: $\quad\quad$ **end for**
13: $\quad$ **end for**
14: **end for**
15: **return** $\mathbf{R}'$

Algorithm 2 Finding End to End Traffic Matrix

### 3. Traffic Pattern Discovery

The traffic matrix R defines the deduced end-to-end traffic volume between each pair of mobile nodes. Even though we still need to perform more investigation to discover the actual source/destination probability distribution and end-to-end link probability distribution, that is, to figure out who are the actual sources and destinations and who are communicating with whom. In traffic pattern discovery system there are three sub modules to discover the traffic, they are source probability distribution, destination probability distribution and end to end link probability distributions. Each sub modules are explained as follows: After finding the traffic matrices using these matrices we have to find the source and destination probability distributions which contain several functions. All nodes within the transmitting range of a packet have the same probability to be the actual sender and receiver. Using vector space similarity assessment two nodes with higher probability to be neighbors have less impact on each other's source/destination probability distribution, which reasonably reduces the neighborhood noise. Finding the actual source using source probability distribution. Also, finding the actual destination using destination probability distribution.

The source probability distribution algorithm is shown below:

**Algorithm 3.** —$Src(\mathbf{R})$.
1: $\mathbf{S_0} = (1/N, 1/N, \ldots, 1/N)$
2: $n = 0$
3: **do**
4: $\quad \mathbf{S_{n+1}} = (\mathbf{\Phi}(\mathbf{R}) \cdot \mathbf{\Phi^T}(\mathbf{R})) \cdot \mathbf{S_n}$
5: $\quad$ normalize $\mathbf{S_{n+1}}$
6: $\quad n = n + 1$
7: **while** $\mathbf{S_n} \neq \mathbf{S_{n-1}}$
8: $\mathbf{S} = \mathbf{S_n}$
9: **return** S

Algorithm 3 Finding Source Probability Distribution

Both vectors which is shown in the algorithm they are uniform probability distribution vectors. Similar approach can be applied for to find the destination probability distributions which are shown in the algorithm4 below. In this all probability distribution vectors are normalized and only the relative orders among the elements of each vector actually make sense. The actual destination can be obtained by using the algorithm4 which is shown below:

**Algorithm 4.** —$Dest(\mathbf{R})$.
1: $\mathbf{D_0} = (1/N, 1/N, \ldots, 1/N)$
2: $n = 0$
3: **do**
4: $\quad \mathbf{D_{n+1}} = (\mathbf{\Phi^T}(\mathbf{R}) \cdot \mathbf{\Phi}(\mathbf{R})) \cdot \mathbf{D_n}$
5: $\quad$ normalize $\mathbf{D_{n+1}}$
6: $\quad n = n + 1$
7: **while** $\mathbf{D_n} \neq \mathbf{D_{n-1}}$
8: $\mathbf{D} = \mathbf{D_n}$
9: **return** D

Algorithm 4 Finding Destination probability distributions

The final part of the traffic pattern discovery system is to find the end to end link between the every node in the network. The procedure and algorithm for finding the end to end link probability or end to end link between the pair of nodes is shown below: In the symmetrical way we have to

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 5, May 2015)

apply the algorithm 5 to find the end to end link between the nodes of destination to source.

Suppress-Sender($i$)
$\mathbf{W}'|_{1\times\mathbf{K}} = \mathbf{W}|_{1\times\mathbf{K}}$
for $p = 1$ to $K$
  for $q = 1$ to $N$
    $w'_p(i,q) = 0$
return $\mathbf{W}'|_{1\times\mathbf{K}}$

Suppress-Receiver($j$)
$\mathbf{W}'|_{1\times\mathbf{K}} = \mathbf{W}|_{1\times\mathbf{K}}$
for $p = 1$ to $K$
  for $q = 1$ to $N$
    $w'_p(q,j) = 0$
return $\mathbf{W}'|_{1\times\mathbf{K}}$

**Algorithm 5.** Given a source node $i$, compute the probability distribution vector $\mathbf{L}_{s-d}(\mathbf{i})$ for each node to be the intended destination of $i$.

1: $\mathbf{R} = f(\mathbf{W}|_{1\times K})$;
2: $\mathbf{D} = Dest(\mathbf{R})$;
3: $\mathbf{W}'|_{1\times\mathbf{K}} = $ Suppress-Sender($i$);
4: $\mathbf{R}' = f(\mathbf{W}'|_{1\times K})$;
5: $\mathbf{D}^- = Dest(\mathbf{R}')$;
6: Calculate the probability reduction vector as: $\mathbf{L}'_{s-d}(\mathbf{i}) = \mathbf{D} - \mathbf{D}^-$. If negative elements exist in $\mathbf{L}'_{s-d}(\mathbf{i})$, increase each element by the absolute value of the smallest negative element;
7: Normalize $\mathbf{L}'_{s-d}(\mathbf{i})$ to generate the probability vector $\mathbf{L}_{s-d}(\mathbf{i})$ for each node to be the intended destination of $i$;
8: Return $\mathbf{L}_{s-d}(\mathbf{i})$.

Algorithm 5 Finding end to end link between the source to destination

## III. EXPERIMENTS AND PERFORMANCE EVALUATION

In the experiment part we have to demonstrate that to find the source, destination and end to end link between the pair of nodes. So for this we have to set up separate demonstration. For this we have to generate three scenarios. After random deployment of 30 mobile nodes in 800*800m2 area with the node speed of 10m/s and transmission rate is of 11Mbps.Now we have to generate the three scenarios which are explained below: In scenario 1 only one source is generates the CBR traffic to four destination .In scenario 2, four source node is generating CBR traffic to only one destination. And finally third scenario is CBR traffic is generated fifteen point to point pair of nodes. After generating this environment we have to simulate each scenario for 200seconds in NS2.34.And we get the probability distributions of source, destination and end to end link between the nodes. In this demonstration part we get the probability versus number of nodes graph. In that graph the node which shows the highest peak level that node is considered as the actual source or destination nodes.

To find the accuracy of this traffic analysis model we have the performance evaluation part. In this we have to increase the number of nodes with specified area. For this we have taken 80 mobile nodes for 1000*1000 m2 area .And same initial parameters are taken which is mentioned in the demonstration section. For performance evaluation we are taken the two strategies as follows: The strategy T1 is suppose the number of actual sources, destinations or end to end links between nodes to be k. At that time we simply select the top k items i.e. nodes or links with the highest probabilities. The second strategy T2 is suppose the number k is unknown. In this strategy also we keep selecting the top items with highest probabilities until two criteria are satisfied: one is, the sum of the probabilities of the selected items has reached u; and other one is probability of the last selected item is v times larger than the current one. Here u and v are the adjustable thresholds. After simulating this we get the performance evaluation factors which are false positive rate and false negative rate. The performance evaluation graphs of source and end to end link identification are shown below: The fig.2 and fig 3 shows the average false positive rate and average false negative rate of source identification respectively. In this, we can see both T1 and T2 strategies achieve reasonably good accuracy for source identification. Using T1 strategy the false positive rate is almost always less than 0.2 and although it increases slightly as the number of sources goes up. And the peak of the false negative rate is lower than the 0.3.T2 tends to select more nodes which inevitably increase the false positive rate at the same time false negative rate is decreases

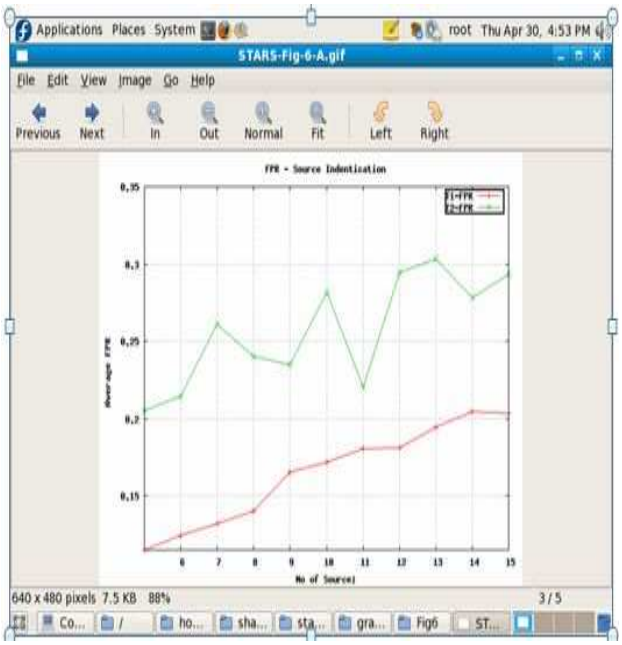


Fig.2 Average Positive Rate of Source Identification

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 5, May 2015)

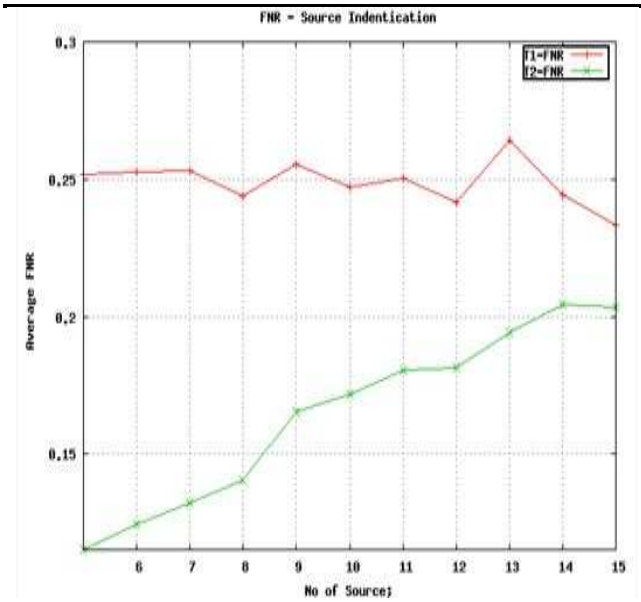


Fig.3 Average False Negative Rate of Source Identification

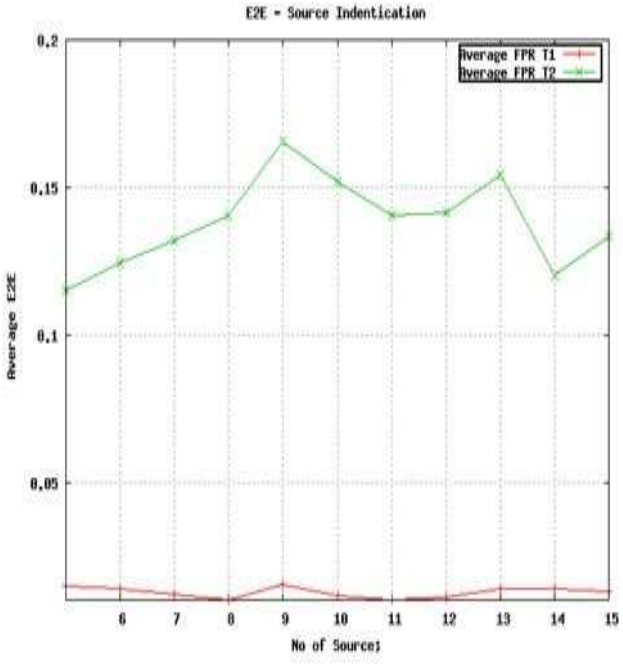


Fig.4 Average False Positive Rate of End to End Link Identification

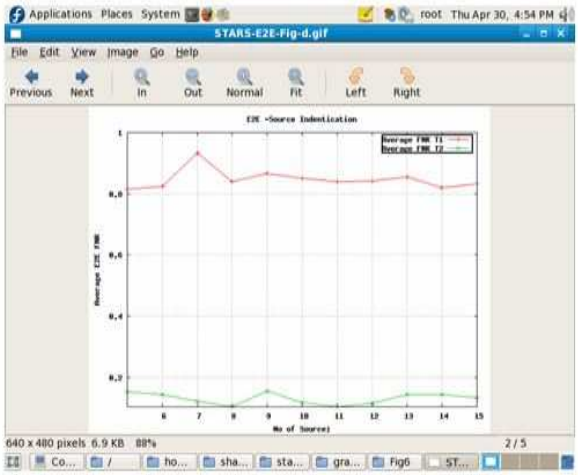


Fig.5 Average False Positive Rate of Source Identification

The fig.4 and fig 5 shows the average false positive rate and average false negative rate for end to end link identification respectively. When the actual end to end links known we got perfect false positive rate which is close to zero but unfortunately the extremely large false negative rate is not acceptable. This explains that too few links are selected and most of the actual end to end communication relations are not discovered. In the second strategy T2 selects more links which reveals most of the actual end to end communication links by slightly sacrificing the false positive rate. In T2 strategy the false negative rate is almost 0.2 and false positive rate is 0.16 which is shown on the above graphs.

Using these performance evaluation graphs, the hidden traffic patterns can be revealed in good accuracy using the statistical traffic pattern discovery system, even without the number of actual sources, destinations and end to end communication relations known to the traffic analyzers.

**Comparison of Proposed traffic analysis model with Existing Traffic analysis Model**

In this section we compare our proposed traffic analysis model with existing traffic analysis model. Here we are using the factors that total packet send, packet drop, average packet delay, pattern detection rate and packet throughput. The details of these comparisons shown in the graphs which are shown below:

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 5, May 2015)

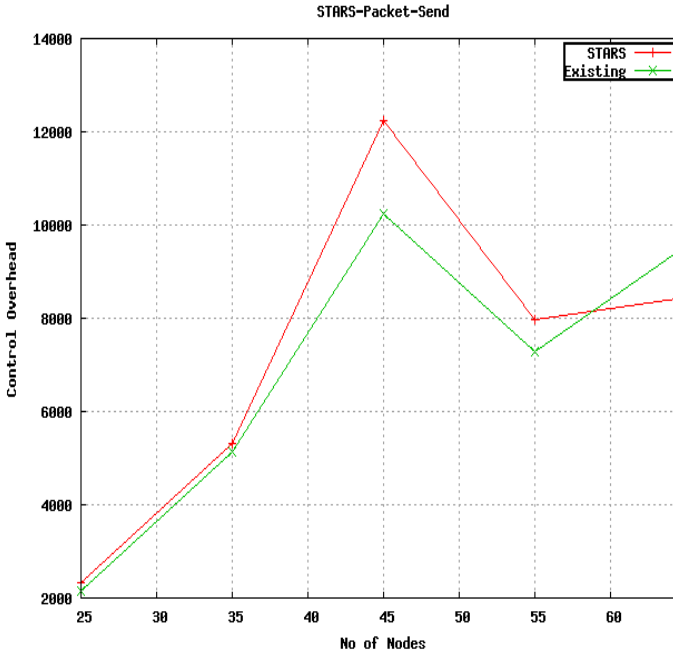


Fig 6.Toatal packet Sending Rate

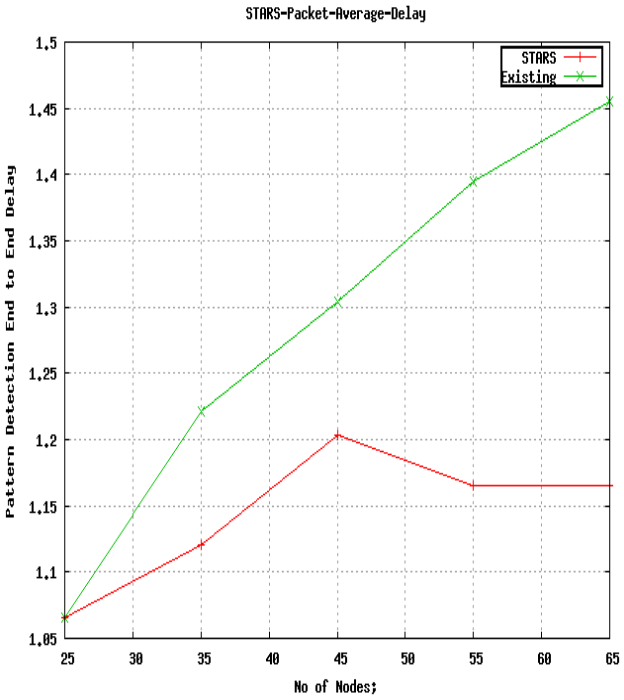


Fig 8.Packet Average Delay

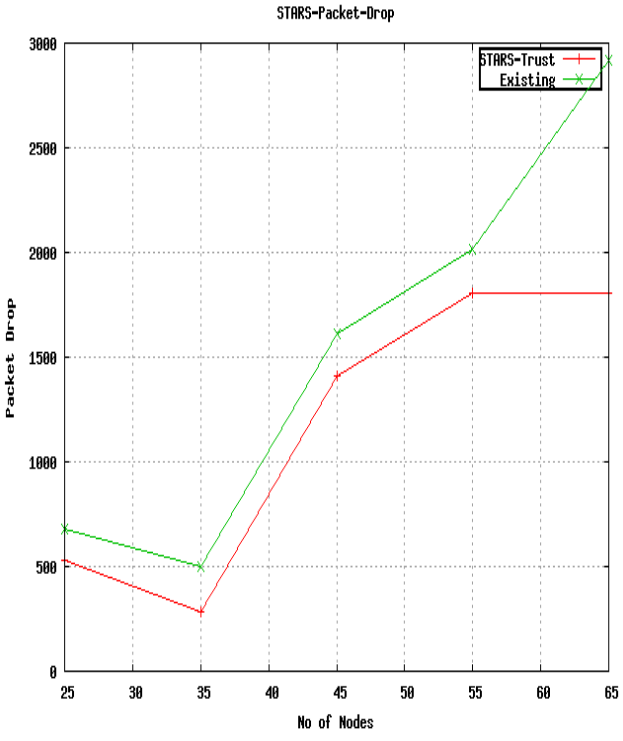


Fig 7.Toatal packet Drop Rate

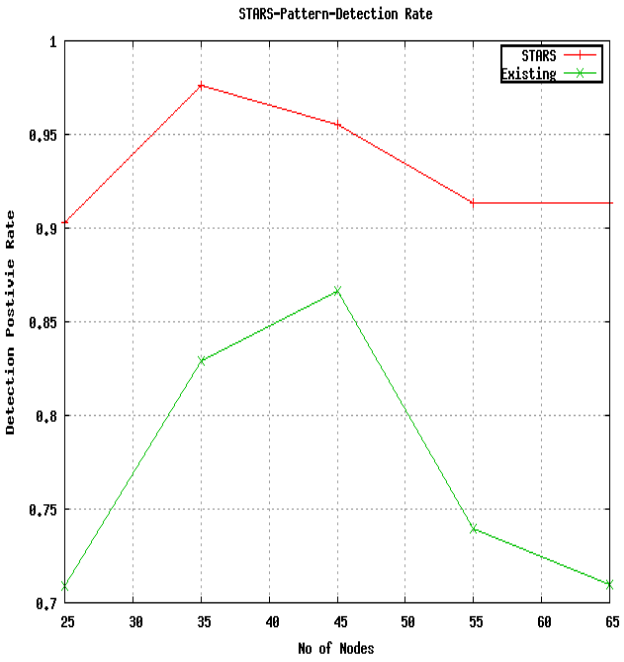


Fig 9.Pattern Detection Rate

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 5, May 2015)

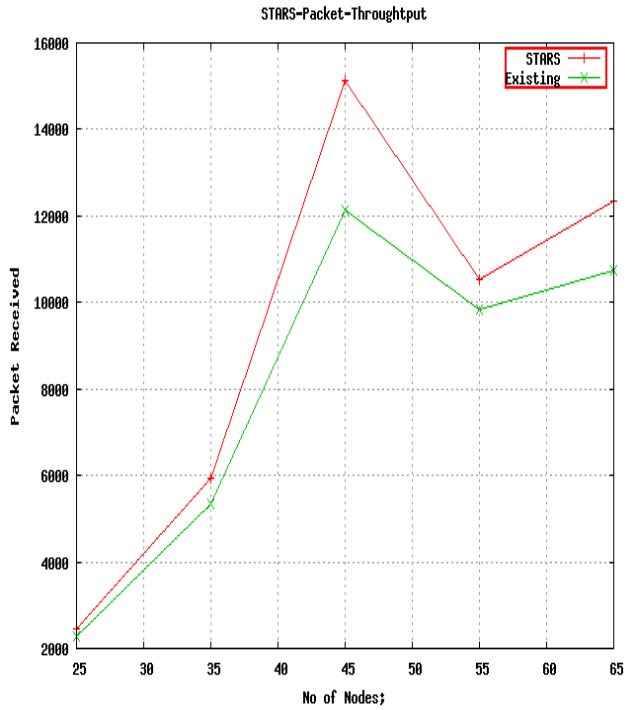


Fig 10.Packet Throughput

In the Fig 6, Fig 7 and Fig 8 shows the comparison details of the packet sending rate, drop rate and packet average delay respectively. Those graphs clearly say the proposed traffic models are better than existing traffic analysis model. In the fig 9 and fig 10 explains the details of pattern detection rate and packet throughput respectively. Those graphs explains the which model is better for pattern detection of anonymous MANETs. Pattern detection rate of proposed traffic analysis model is better than existing model which is clearly explained in the fig 9.Most important graph is fig 9 because the aim of our proposed model is proved in that graph i.e. which shows the proposed traffic analysis model's pattern detection rate is better than existing traffic analysis model.

## IV. CONCLUSION

In this traffic analysis model we can reveal the hidden traffic patterns of anonymous MANETs without decrypting the packets. The hidden traffic patterns can be revealed in good accuracy using the statistical traffic pattern analysis model, even without the number of actual sources, destinations and end to end communication relations known to the traffic analyzers. Also comparisons graphs of performance evaluation section and comparisons section proves that our proposed statistical traffic analysis model shows better accuracy for pattern detection in anonymous MANETs. Finally the proposed traffic analysis model is detects the 80% of statistical traffic patterns of anonymous mobile ad hoc networks.

## REFERENCES

[1]. 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. Zhang, W. Liu, W. Lou, and Y. Fang, "MASK: Anonymous On-Demand Routing in Mobile Ad Hoc Networks," IEEE Trans.Wireless Comm., vol. 5, no. 9, pp. 2376-2385, Sept. 2006.

[3]. Y. Qin and D. Huang, "OLAR: On-Demand Lightweight Anonymous Routing in MANETs," Proc. Fourth Int'l Conf. Mobile Computing and Ubiquitous Networking (ICMU '08), pp. 72-79, 2008.

[4]. D. Huang, "Unlinkability Measure for IEEE 802.11 Based MANETs," IEEE Trans. Wireless Comm., vol. 7, no. 3, pp. 1025- 1034, Mar. 2008.

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

[6]. Yang Qin, Dijiang Huang, "STARS: A Statistical Traffic Pattern Discovery System For Anonymous MANETs" IEEE Transactions On Dependable and Secure Computing, vol. 11, No. 2, March/April 2014

[7]. 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.

[8]. G. Danezis, "Statistical Disclosure Attacks: Traffic Confirmation in Open Environments," Proc. Security and Privacy in the Age of Uncertainty (SEC '03), vol. 122, pp. 421-426, 2003.

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

[10]. G. Danezis, C. Diaz, and C. Troncoso, "Two-Sided Statistical Disclosure Attack," Proc. Seventh Int'l Conf. Privacy Enhancing Technologies, pp. 30-44, 2007.

[11]. 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 (SECON '10), pp. 1-9, 2010.

[12]. A. Boukerche, K. El-Khatib, L. Xu, and L. Korba, "SDAR: A Secure Distributed Anonymous Routing Protocol for Wireless and Mobile Ad Hoc Networks," Proc. IEEE 29th Ann. Int'l Conf. Local Computer Networks (LCN '04), pp. 618-624, 2004.

[13]. S. Seys and B. Preneel, "ARM: Anonymous Routing Protocol for Mobile Ad Hoc Networks," Proc. IEEE 20th Int'l Conf. Advanced Information Networking and Applications Workshops (AINA Workshops '06), pp. 133-137, 2006.

[14]. R. Shokri, M. Yabandeh, and N. Yazdani, "Anonymous Routing in MANET Using Random Identifiers," Proc. Sixth Int'l Conf. Networking (ICN '07), p. 2, 2007.

[15]. R. Song, L. Korba, and G. Yee, "AnonDSR: Efficient Anonymous Dynamic Source Routing for Mobile Ad-Hoc Networks," Proc. Third ACM Workshop Security of Ad Hoc and Sensor Networks (SASN '05), pp. 33-42, 2005.

[16]. 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