Comprehensive Study of Traffic Analysis In MANET

T. Parameswaran[1], Dr. C. Palanisamy[2], M.Karthigadevi[3]

T. Parameswaran is with Assistant Professor, Department of Computer Science and Engineering, Anna University Regional Centre, and Coimbatore, India; (e-mail: tparameshse@gmail.com).

Dr.C.Palanisamy is with Professor and Head, Department of Information Technology,Bannari Amman Institute Of Technology, Sathyamangalam, India; (e-mail: cp samy@yahoo.com).

M.Karthigadevi is with PG Scholar, Department of Computer Science and Engineering,Anna University Regional Centre, Coimbatore, India; (e-mail: karthigamurugan29@gmail.com).

Abstract: Mobile Ad hoc Networks (MANETs) are in dynamic in nature and have a wireless radio medium with limited resources and lack of centralized administration. So, Anonymous communication is the big challenge in MANETs. It is very hard to observe the source node, destination node, and their relations between them in any given network. Many techniques are proposed to enhance the anonymous communication. However, MANETs are vulnerable under traffic analysis. This paper describes the traffic analysis, explain some of the attacks under traffic analysis, how they could infer on wired networks and MANETs.

Keywords: Mobile Ad hoc Networks, Anonymous communication, Traffic Analysis.

1. INTRODUCTION

MANETs are an infrastructure free, wireless and self configurable network for mobile devices. They are used in military missions. If nodes are present within their transmission range they will use point-to-point communication which means two or more mobile nodes can communicate each other directly in MANETs. Otherwise they will use intermediate nodes between the source and destination to deliver the packets (end-to end communications). Here there is no centralized administrator among the nodes. Hence, different nodes maintain the routing and the resource management in distributed manner. So MANETs is vulnerable under traffic analysis. It is difficult to track the destination of the source node in the communication link and other intermediate nodes involved in it. And also finding the valuable information (Start time, end time, rate, etc.,) is difficult. Anonymous routing protocol (ANODR, MASK, OLAR) aims to prevent inferring the traffic pattern by hiding the real sources, destination and their relationships of the overall packets which encourages anonymous MANET communication. The following methods are used for anonymous MANET communication. 1) Route discovery: proactive or reactive (on demand or table driven routing). 2) Data delivery: source routing by multilayer encryptions or end-to-end re-encryptions by a virtual circuit. 3) Trust management: including both key management and identify/pseudonym management. 4) Modifying control functions: masking MAC/IP addresses, disabling virtual carrier sensing, and discarding ACK packets. Through these methods, anonymous MANET communication can provide: (i) Anonymity of sender/receiver – that is critical to detect the sender and receiver node of captured packets. (ii) Anonymity of communication relating – That is critical to identify the end-to-end communication relations between mobile nodes.

However, we can still track the routing information via the traffic analysis. In this paper, our prime focus will be on some of the traffic analysis. This paper is organized as follows: section2 describes the general traffic analysis and how they used in earlier approaches. Section3 presents the statistical traffic analysis. Section4 describes types of passive attacks used in MANETs. Section5 presents conclusion of this paper.

2. TRAFFIC ANALYSIS

Traffic analysis is the process of intercepting and collecting messages on the way to track important information from patterns in communication. That information is not leaking or not modified just do monitoring and analysis activities. It can get information from the monitor frequency and timing packets. This analysis is not only used for collecting information and also used to bypass security mechanisms in place. It can perform encryption only not decryption. Traffic analysis can be applied in military and counter applications.

In the earlier approaches, traffic analysis models have been widely used for static wired networks. For example, the simplest approach to observe a message is to capture them one by one all possible path a message could traverse, namely the brute force method [1]. Node flush attack, Timing attack [1] also an earlier approaches for traffic analysis.

2.1 Brute force attack

Brute force attack [1] is very instructive due to it helps in determining how much, when and where to use dummy traffic. These dummy messages are messages that are sent through the network that complicate the adversary's process. We present these adversary in a setting in which each mix node waits until it receives t messages before flushing them. In addition, we assume each message goes through exactly do mix nodes.

(1) The adversaries first follow a message from a source to a first mix node.
(2) Then the adversaries follow every message that the first node releases. The adversary only needs to monitor one node, if all the messages send to either the same mix node or receivers. If all t messages sent to different nodes, the adversaries required to analyze t different mix nodes.
(3) That process continues in this manner until messages reach the death level nodes. The adversaries, then need only "follow" messages.

The brute force attack can be carried out by passive, static, external attackers and can be applied in only wired networks.

2.2 Node flushing attacks

The flush attack [1] is very effective and mounted by an active external attacker. If the nodes wait until they have t messages before "flushing", an adversary can send it-1 messages and easily associate messages leaving the node with those having entered. That can be viewed by noting that the observer will be able to match his inputs with the messages leaving the node. The observer can't distinguish the dummy traffic from legitimate messages. In specific instances dummy traffic can't be used. Authenticating each message and detecting flushing attempts which could be computationally infeasible. The adversaries can take these advantages. But these attacks only suitable for wired networks.

2.3 Timing attacks

The system could be vulnerable to timing attacks because different routes can be taken different amounts of time [1]. When the network uses mix nodes the attack is very effective. Mix nodes with a variable amount of time before flushing messages. This attack can be carried out by passive observer and only suitable for wired networks.

3. STATISTICAL TRAFFIC ANALYSIS

The statistical traffic analysis attack defines the network information from its statistical characteristics which are different from above mentioned passive attacks. In these attacks the adversary does not modify the traffic behavior either by modifying or deleting the data packets. They just gather the packets, then do the statistical analysis. Predecessor attack, disclosure attack are some of the examples of statistical traffic analysis.

3.1 Predecessor attacks

In Predecessor attack [2], the adversaries detect an identifiable stream of communications over a number of rounds. The adversaries simply log any node that sends the message in each round. The attack must require to exploit the process of path initialization rather than timing analysis and packet size. The Predecessor attack works in the following manner.

- The subset of nodes that forward a source messages is selected uniformly at random.
- The source makes repeated connections to specific destinations, which are made outside the group performing the protocol.

In order to find time analysis required for the attack and that source do not leave from the group until the adversaries are successful. These are needed for the attack work in all cases, and they are also critical to the process of anonymous routing protocol. Crowds, Onion routing, Mix net, DC-net, etc., are some of the anonymous routing protocols they are used to detect and destroy the predecessor attacks.

Crowds: It uses a collection of nodes that act as proxies for a given initiation from the group. An initiation message is forwarded from the source to a series of proxy, forming a path for all future information from the source. Based on this receiving message, each proxy decides, based on a probability of forwarding whether to extend the path through another proxy chose at random with uniform probability or to become the last node on the path and directly communicate with the destination. That path is managed for a limited period, after which all the path must be reformed. The time limit allows nodes that merge the protocol to add path at the same time as all other nodes else new paths may be easily assigned to recently merged nodes.

Onion Routing: Onion routing is similar to crowds in which an initial message forms a path of proxies through which the source sends its future messages. The protocol gets name from that method of encrypting the initial packet and address of the proxies at each hop in the path with the public key of the previous step. This method results in layers of encryption that are peeled off at each step. In this order we have to find the next address to send to on the path. That needs the source to predetermine the entire path.

Mix-net: Some of the protocols for anonymity communications are web mixes, stop and go mixes, onion routing, etc., Mix-net protocol that uses onion-routing's layered encryption and also supports mixing techniques for timing analysis. These techniques include sending information in reordered batches, sending dummy information and introducing random delays. Mix-net protocol stops timing analysis effectively.

P5: P5 is designed for anonymity between peers to connect each other. It uses tree-based broadcast protocol where a user's anonymity is based on the size of the different broadcast groups. When the number of attackers greater than the size of user's anonymity group.

Tarzar: Tarzar is a peer to peer system at the network layer. When the number of honest domains represented in the TARZAR group is small, the adversaries can gain an advantage. With no participants the adversaries may be able to operate nodes. The adversaries in that domain could make it more likely to appear on a source path despite only operating a limited corrupt nodes.

Morph-Mix: Morph Mix is also a peer-to-peer paths based protocol. It allows honest participants to detect adversaries in the system over time. Here the peers do not need to know all other peers in the network to operate correctly.

High latency communication systems need messages to be anonymized should be forwarded through a sequence of intermediary nodes which nodes called mixes. Mixes hide the correspondence between input and output messages. That deployed system makes use of pool mixes. These mixes collects a number of messages in each round change their appearance and locate them on a pool from that they are probabilistically chosen to leave the mix. Otherwise, messages stay in the pool, get mixed messages enter in subsequent rounds. The following attacks can act against mix systems.

3.2 Traffic Confirmation Attacks

In Traffic Confirmation attacks [4] all the nodes should act as mixes, subsequent systems built and deployed make a distinction between clients and mix nodes from its core. That distinction is understandable by an attacker that sets as his goal to detect the specific recipient of messages injected into the network or detect back the originators of messages coming out of the network. The attackers try to relate the source and destinations using information present at the mix network where messages are injected or rejected. This attack does not rely on track messages through the network.

3.3 Intersection Attacks

Another family of attacks against mix systems is intersection attack. In intersection attack [4] various messages utilize the same route through the network to perform traffic analysis. It is applied to an entire anonymity system. It does not rely upon any particular properties of mixing other than the unlinkablity it provides.

3.4 Hitting Set Attack

Hitting set attack [5] needs minimum observations by looking for unique minimal hitting sets. This attack is enhanced by the use of frequency analysis that is possible to use highly efficient backtracking search algorithm.

Statistical minimal hitting set attack needs less observation than any other attacks, however, it requires the solution of NP-hard problem. It can be applied in any situations that were far beyond feasibility of the attack and it is easy to scalable. Statistical Hitting set attacks find approximate solutions which lead to possible errors. The error can be reduced by applying some error probability, statistics and strategies. This attack, identifies the recipient set unambiguously.

3.5 Disclosure attack

Disclosure attack [3] observes a mix based system long time enough can uncover a persistent communication pattern. This attack relies on graph theory to uncover the recipient set of a target user. Disclosure attack is simple model where single mix is used by b participants each round, one of them always being an end user, when the other chosen randomly out of a total number of N-1 possible ones. They try to identify mutually disjoint sets of recipients from the sequence of recipient anonymity sets corresponding to end users' messages. Since, it takes time exponential in the number of messages to be analyzed that operation is the bottleneck for the adversaries. The main drawback is its reliance on solving a constraint satisfaction problem which is called NP-complete. This is applied in wired networks due to discovering the sender is critical in case of MANETs.

3.6 Statistical disclosure attacks

Statistical Disclosure attack [4] can try to overcome the problem in Disclosure attack. This attack is possible that yields the set of potential recipients of the source. That set can be used to detect the recipients of a particular message sent out by the end user. It can be applied when the probability distribution described by vectors are not uniform but are skewed. This attack defines vector with N elements to each potential recipient of a message in the system. The probability distribution is used by source to choose the recipient of its messages for each round of the abstract mixing. This attack is computationally cheap and scales very well because it only relies on collecting observations and performs trivial operations on vectors. In statistical disclosure attack Computational improvement is simply not comparable to disclosure attacks, but also present new features. The signal detection problem is to differentiate the signal source from the noise introduced by other senders. The statistical disclosure attack has conditions which can be expressed in the closed algebraic method. Without simulation, we have to decide when it is applicable and effective. The recipient anonymity set of a message is detected with respect to the result of that assignment instead of assigning uniform probabilities among all recipients as SDA does.

3.7 Perfect matching disclosure attacks (PMDA)

In previous approaches on disclosure attack [6], mixes provides anonymity communication is broken by specific user behavior. Previous work is considered a very simple model, where users and messages to a fixed set of contacts through a threshold mix. Here users use the uniform probability to choose their communication partners.

The perfect matching disclosure attack [6] is an efficient attack based on graph theory and requires without any assumption on the user behavior in order to identify their relationships. In previous disclosure attack sender sends exactly one message per round. It is more effective when de-anonymizing mixing rounds due to consider all nodes at once in a round rather than single nodes iteratively. Here chose a simple threshold, mix as a communication channel for both Perfect Matching Disclosure attack and Statistical Disclosure attack. Both perform very similar action with respect to the simple user behavior model. The accuracy is good in PMDA compared to statistical disclosure attack. This attack detects the relationship between sending and received messages in a round that is one-to-one to improve the accuracy of the evaluated profiles. PMDA notes this interdependency by detecting for perfect matching in the bipartite graph representing mix round, when NSDA normalizes the adjacency matrix representing that graph.

4. STATISTICAL TRAFFIC ANALYSIS IN MANETS

The Predecessor attack and Disclosure attack are not suitable for MANETs to effectively analyze the traffic because of the following nature of MANETs. They are: i) the broadcast nature – It is critical to detect the destination exactly because where the packets sent and received by many nodes. ii) The ad hoc nature-the ad hoc networks are infrastructure free so that every node act as both ender and receivers. Hence it is critical to detect the nature of the node. iii) The mobile nature-here the communication between nodes are very difficult to analyze because nodes change their location.

4.1 Evidence based statistical traffic analysis

In evidence based statistical traffic analysis [10] each data packets are captured which are considered as evidence that support a point to point transmission between sender and receiver. In this analysis first create a sequence of point-to-point matrices, and then using that matrices derive end-to-end relations between the communication paths. This method fails when deriving the multi-hop traffic from the one hop evidences. This approach does not provide any method to detect the actual source and destination. It utilizes a naïve accumulative traffic ratio to detect the multi hop communication which leads a lot of inaccuracies in the derived probability distributions.

4.2 Traffic inference in anonymous MANETs (TIA)

Anonymous routing protocol aims to prevent observing the traffic pattern by hiding the real senders, receivers and their relationships of the overheard packets. But using Traffic Inference algorithm [12], a global adversary can observe the MANET traffic pattern accurately. The algorithm assumed the relation between data frames, routing frames, and MAC frames enable to the passive observer, which permits the observers to detect the single-hop traffic using MAC frames, thereby allows to find the multi-hop traffic using routing frames and finally traces the traffic pattern using data frame. This algorithm exploits the overheard routing frames for flow recognition and then detect each information flow in rounds depends upon the data frame inter arrival times. It derives the hidden flows by visible the adversary and thus the traffic pattern from overheard MAC frames without prior knowledge of the inter arrival time distribution of any flow. It works on low latency mix networks. The passive observer divides time into periods and uses TIA to trace the traffic pattern in each period. The consecutive period traffic pattern aggregated to obtain the long term pattern or analysis to observe the pattern changes. TIA consists of three key components:

- *Evidence generation:* The attacker partitions all the MAC frames overheard at the end time into data frames, MAC control frames and routing frames.
- *Flow recognition:* The attacker recognizes each flow except its traffic volume and target period by analyzing routing frames.
- *Traffic Inference:* The attacker decides the traffic volume and end time of each flow by analyzing data frames and thus detects the traffic pattern.

In inter arrival-based algorithm a passive global adversary can detect the traffic pattern accurately through the use of anonymous on-demand MANET routing protocols. The simulation result shows TIA can infer the traffic pattern with an accuracy as high as 95%.

4.3 Statistial traffic pattern discovery system (STARS)

The Statistical Disclosure Traffic Pattern Discovery System [11] uses the concept of heuristic approach is used to find the hidden traffic pattern in MANETs. It performs the traffic analysis based on statistical nature of captured raw traffic. Using this method the passive observer observes the actual source and destination nodes, and then correlates the source to their corresponding destination. It reused the evidence-based model and then derived the source/destination probability distribution and multi-hop probability distribution used to detect the traffic pattern. All previous methods are used for partial attack, they can’t detect both the source and destination at the same time for any given network. This attacking system detects all source and destinations and also traces their relationship between them.

4.4 Least square disclosure attack (LSDA)

LSDA [7], [8] ensures the error between the actual number of messages each user gets from the mix and prediction messages sent to the mix is minimized. It takes the advantage of the deterministic routing behavior of the position based routing. The traffic disclosure problem modeled as an efficiently solvable and linearly constrained least square problem.

A mobile ad hoc network protected by anonymity enhancing techniques such that all information flows are encrypted. The traffic analyzer cannot decrypt the information flows, and neither can they reveal the multi-hop communication relations from the routing layer and above. However, the adversaries capture every packet transmitted in the network. Using location tracking systems, adversaries are aware of the locations of the mobile nodes at any given time. The MANET uses a position based routing strategy whose routing assumptions are reproducible given the nodes' positions. In addition, the traffic distribution is given by STARS or similar traffic profiling schemes in the network, what is called the profiled traffic distribution. The main aim is to deanonymize the network communications on a per-flow basis. The previous works estimated the attacks either from mostly de-anonymization of individual messages or from the point of view number of rounds needs to find a percentage of end users.

The Least Square Disclosure attack [14] is an effective estimator when user's behavior is static that attack is suitable against anonymous communication through both threshold and pool mixes. Here messages are individually selected, to stay in the mix or to be sent to receiver according to a binomial distribution in threshold binomial pool mixes. The paper takes [8] two variants of LSDA: A very effective unconstrained profile estimator which output user profiles contain probability that may be negative. Slower constrained profile that minimizes the error through ensuring the output profiles are well defined. The feature of this approach is that it permits for the derivation of analytical expressions that describe the estimation or the error profiling with the system parameters. It allows designers to select system parameters that provide a certain level of protection no need to run simulations. In this system model we consider two types of mixes:

(1) *Threshold Mix:* This mix collects threshold messages per round, transforms cryptographically and outputs them in a random order, then, hiding the correspondence between incoming and outgoing messages**.**

(2) *Binomial Threshold Pool Mix:* This mix gathers threshold messages each round and change their appearance to ignore bitwise linkability. However messages are placed in a pool and only leave the mix with the firing probability of the binomial pool mix. Else, they stay and receives mixed with messages arriving in subsequent rounds.

Table 1: Summary of types of statistical traffic analysis

Types of Attack	**Works on**	**Procedure**	**Advantages**	**Disadvantages**
Predecessor Attack	Wired networks	It maintains the counter to interact with other nodes in the communication network, which is used to observe traffic information.	It just collects packets without changing the network behavior.	Many routing protocols used against this attack. It's not suitable for MANET because of that nature.
Disclosure	Wired	It observes a mix based system using	It identifies the set	Solving NP-

attack	networks	graph theory. It traces mutually disjoint set of receivers among the recipient anonymity sets of messages sent by the end user.	of end user contacts	complete problem
Statistical Disclosure Attack	Wired networks	It can isolate end user behavior by evaluating the behavior of other users, managing those observations Where end user not participated.	It is a more general scenario used for more complex mixing algorithms.	High computational complexity.
Perfect Matching Disclosure Attack	Wired networks	It can be applied in simple threshold mix. It considers all nodes at once in a round rather than single nodes iteratively in order to identify their relationships using graph theory.	No need of assumption about the user behavior	More expensive in computational complexity.
Hitting set Attack	Wired networks	This attack is done by frequency analysis. It requires less observation rather than any other attacks.	It speeds up the search for end user's recipients by denying the search to unique minimal hitting set.	It provides approximate solutions which lead risk of errors.
Evidence based statistical traffic Analysis	MANET	Here first create a sequence of point-to-point matrices, and then using that matrices derive end-to-end relations between the communication paths.	It is a good practice attacking system against MANET	i) Not suitable for multi-hop traffic. ii) Lot of inaccuracies
Statistical Traffic Pattern Discovery System	MANET	Using this method the passive observer observes the actual source and destination nodes, and then correlates the source to their corresponding destination using heuristic approach.	It determines the source, destination and their relations effectively.	i) It can't deanonymize communication on a flow basis. ii) It can't take the special features of particular traffic.
Least Square Disclosure Attack	MANET	It is an effective estimator when user's behavior is static that attack is suitable against anonymous communication through both threshold and pool mixes. It detects the error between the actual number of messages each user gets from the mix and prediction messages	It can be deanonymize communication on a flow basis. It can be applied in high latency anonymous	It performs partial attacks.

		sent to the mix is minimized.	communication system such that both threshold and pool mixes.	
Traffic Inference Algorithm	MANET	This algorithm assumed the relation between data frames, routing frames, and MAC frames enable to the passive observer, which permits the observers to detect the single-hop traffic using MAC frames, thereby allows to find the multi-hop traffic using routing frames and finally traces the traffic pattern using data frame.	It infers the traffic pattern with an accuracy as high as 95%.	It works on low latency mix networks such that on demand routing protocols.

5. CONCLUSION

In this paper, we present an overview of Traffic analysis and some of the previous approaches used in traffic analysis how they are working in the system. We investigated types of statistical traffic analysis used in both wired networks and MANETs and their advantages and disadvantages are discussed. Compared to Statistical Traffic Pattern Discovery System other attacking system can perform partial attacks only that they either only try to detect the source or to determine the corresponding destination. STARS first detects all source and destination and then find out their communication relationship. But STARS can be applied to limited nodes.

REFERENCES

[1] Raymond (2001): Traffic analysis: Protocols, attacks, design issues, and open problems. Lecture Notes in Computer Science, pp. 10–29.

[2] Wright, K.; Adler, M.; Levine, B. and Shields, C. (2002): The Predecessor Attack: An Analysis of a Threat to Anonymous Communications Systems. Computer Science Department Faculty Publication Series. Paper 164.

[3] Agrawal, D. and Kesdogan, D. (2003): Measuring anonymity: The disclosure attack. IEEE Security & Privacy, vol. 1, no. 6, pp. 27–34.

[4] Danezis, G. (2003): Statistical disclosure attacks: Traffic confirmation in open environments. Security and Privacy in the Age of Uncertainty, D. Gritzalis, P. Samarati, and S. K. Katsikas, Eds. Athens, Greece: Kluwe, pp. 421–426.

[5] Kesdogan, D. and Pimenidis, L. (2004): The hitting set attack on anonymity protocols. Information Hiding (Lecture Notes in Computer Science), vol. 3200, J. J. Fridrich, Ed. New York, NY, USA: Springer-Verlag, pp. 326–339.

[6] Troncoso, C.; Gierlichs, B.; Preneel, B. and Verbauwhede, I. (2008): Perfect matching disclosure attacks. Privacy Enhancing Technologies (Lecture Notes in Computer Science), vol. 5134, N. Borisov and I. Goldberg, Eds. New York, NY, USA: Springer-Verlag, pp. 2–23

[7] Pérez-González, F.; Troncoso, C. (2012): A least squares approach to user profiling in pool mix-based anonymous communication systems. Proc. IEEE Workshop Inform. Forensics Security, pp. 115–120.

[8] Perez-Gonzalez, F.;Troncoso, C. (2012): Understanding Statistical Disclosure: A Least Squares approach. Privacy Enhancing Technologies Symposium (M. Wright and S. Fischer-Hubner, eds.), vol. 7384 of LNCS, pp. 38–57, Springer-Verlag.

[9] Mathewson, N.; Dingledine, R. (2004): Practical traffic analysis: Extending and resisting statistical disclosure. Privacy Enhancing Technologies (Lecture Notes in Computer Science), vol. 3424, Springer-Verlag, pp. 17–34.

[10] Huang, D. (2008): Unlinkability Measure for IEEE 802.11 Based MANETs. IEEE Trans. Wireless Comm., vol. 7, no. 3, pp. 1025-1034.

[11] Yang Qin, Dijiang Huang and Bing Li (2014): STARS: A Statistical Traffic Pattern Discovery System for MANETs. IEEE Transaction On Dependable And Secure Computing, Vol. **11, No.** 2.

[12] Liu, Y.; Zhang, R.; Shi, J. and Zhang, Y. (2010): Traffic inference in anonymous MANETs. Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 7th Annual IEEE Communications Society Conference, pp. 1–9.

[13] He, T.; Wong, H. and Lee, K. (2008): Traffic analysis in anonymous MANETs. Military Communications Conference, .MILCOM 2008. IEEE, pp. 1–7.

[14] Pérez-González, F.; Troncoso, C. and Simon Oya (2014): A Least Squares Approach to Static Traffic Analysis of High Latency Anonymous Communication Systems. IEEE Transactions On Information Forensics And Security, Vol. 9, No.9.