

Preserving Traffic Privacy in Wireless Mesh Networks

Taojun Wu, Yuan Xue and Yi Cui *

1 Introduction and Motivation

Recently, multi-hop wireless mesh network (WMN) has attracted increasing attention and deployment as a low-cost approach to provide last-mile broadband Internet access [4]. However, to further widen the deployment of WMN, privacy issue must be addressed. For example, in a community mesh network, the traffic of a residence can be observed by the mesh routers residing at its neighbors, which could reveal sensitive personal information. Despite the necessity, limited research has been conducted towards privacy preservation in WMN. This motivates us to investigate the privacy preserving mechanism in WMN.

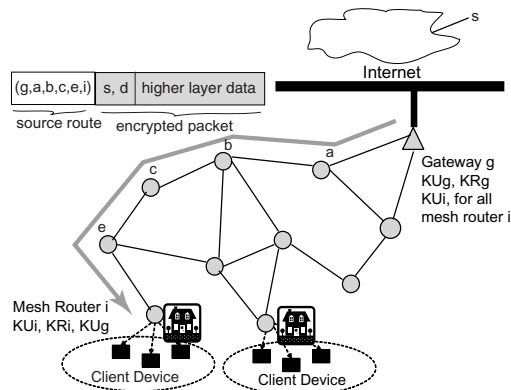


Figure 1. Privacy preserving architecture for wireless mesh network.

We consider a multi-hop WMN shown in Fig. 1. In this mesh network, client devices access a stationary wireless mesh router at its residence. Multiple mesh routers communicate with one another to form a multi-hop wireless backbone that forwards user traffic to the gateway which is connected to the Internet. Our mesh network privacy preserving architecture targets two privacy issues: (1) *Data confidentiality* aims to protect the data content from eavesdropping by the intermediate mesh routers using cryptography-based approach; (2) *Traffic confidentiality* prevents the traffic analysis attack from the mesh routers, which aims at deducing the traffic information such as who the user is communicating with, the amount and time of traffic. In this paper

*Taojun Wu, Yuan Xue, Yi Cui are affiliated with the Department of Electrical Engineering and Computer Science, Vanderbilt University. Their email addresses are {taojun.wu, yuan.xue, yi.cui}@vanderbilt.edu.

we focus on traffic confidentiality, and study the problem of traffic pattern concealment via routing control. In the existing literature, anonymous overlay routing [3, 6, 5] and traffic padding [7] have been proposed to preserve user traffic privacy and increase the difficulty for traffic analysis. In wireless ad-hoc networks, [8, 2] proposed schemes for location and identity privacy. However none of them can be applied to WMN directly. First, the number of nodes in a WMN is limited. Second, traffic forwarding relationship among nodes is strongly dependent on their locations and the network topology, which is static and known a priori in WMN. To better utilize the wireless channel resource and enhance the data delivery performance, a short path is usually selected. Such observations show that the anonymity systems, which rely on relaying traffic among nodes (randomly selected out of thousands) to gain anonymity, can not effectively preserve users' privacy in WMN, or at the cost of significant performance degradation. On the other hand, traffic padding mechanism consumes a considerable amount of network bandwidth, which makes it impractical in resource-constrained WMNs.

In light of these problems, we aim at designing a light-weight traffic privacy preserving mechanism for WMN which is able to balance between traffic analysis resistance and bandwidth cost. Our mechanism makes use of the intrinsic redundancy of WMN, which enables multiple paths for data delivery. We follow the intuition that, if the traffic from the source (*i.e.*, gateway) to the destination (*i.e.*, mesh router) is split to many paths, all the relaying nodes along the paths could only observe a portion of the entire traffic. Moreover, if the traffic is split in a random way both spatially and temporally, an intermediate node has limited knowledge to figure out the overall traffic pattern. Thus the traffic pattern is concealed. Based on this intuition, we seek a routing scheme such that the statistical distributions of the traffic observed at intermediate relaying nodes are independent of the actual traffic from source to destination. To achieve this goal, we first define an information-theoretic metric – “*traffic entropy*” in the following section.

2 Model and Traffic Entropy

We model the WMN shown in Fig. 1 as a graph $G = \{\mathcal{V}, \mathcal{E}\}$, where \mathcal{V} is the set of wireless nodes in WMN, and

\mathcal{E} is the set of wireless edges (x, y) between any two nodes x, y . Each node x maintains a logical connection with the gateway node g . Node x receives data from the Internet via g . The source and destination information of a packet is open to the relaying nodes.

If the traffic between s and x goes through only one route, any relaying node y on this route can easily observe the entire traffic between g and x , thus violating its traffic pattern privacy. To avoid this, x must establish multiple paths with g and distribute its traffic along these paths in a time-variant fashion, such that the traffic pattern observed at any node y is statistically deviant from the original x .

We propose to use traffic entropy as the metric to quantify the performance of a solution at preserving the traffic pattern confidentiality.

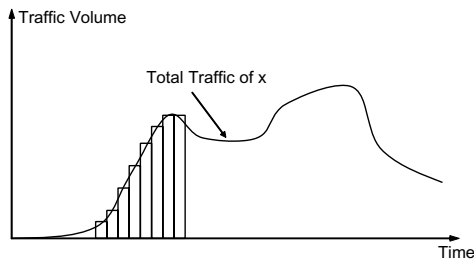


Figure 2. Sampling-based traffic analysis

Ideally, we view the traffic of x as a continuous function of time, as shown in Fig. 2. In practice, the traffic analysis is conducted by dividing time into equal-sized sampling periods, then measuring the amount of traffic in each period, usually in terms of number of packets, assuming the packet sizes are all equal. Therefore, as the first step, we discretize the continuous traffic curve into piece-wise approximation of discrete values.

Now, we use X as the random variable of this discrete value. Y^X is the random variable representing the number of packets destined to x observed at node y in a sampling period. We denote $P(X = i)$ as the probability that the random variable X is equal to i (node x receives i packets in a sampling period). Likewise, $P(Y^X = j)$ is the probability that Y^X is equal to j . ($i, j \in \mathcal{N}$)

Then the discrete Shannon entropy of the discrete random variable X is

$$H(X) = - \sum_i P(X = i) \log_2 P(X = i) \quad (1)$$

$H(X)$ is a measurement of the uncertainty about outcome of X . $H(X)$ takes its maximum value when the value of X is uniformly distributed. On the other hand, if the traffic pattern is CBR (Constant Bit Rate), then $H(X) = 0$, since the number of packets at any sampling period is fixed.

Similarly, we have the entropy for Y^X as follows.

$$H(Y^X) = - \sum_j P(Y^X = j) \log_2 P(Y^X = j) \quad (2)$$

We then define the conditional entropy of random variable Y^X with respect to X as

$$H(X|Y^X) = - \sum_j P(Y^X = j) \sum_i p_{ij} \log_2 p_{ij} \quad (3)$$

where $p_{ij} = P(X = i|Y^X = j)$ is the probability that $X = i$ given condition that $Y^X = j$. $H(X|Y^X)$ can be thought of as the uncertainty remaining about X after Y^X is known. The joint entropy of X and Y^X can be shown as

$$H(X, Y^X) = H(Y^X) + H(X|Y^X) \quad (4)$$

Finally, we define the mutual information of X and Y^X as

$$\begin{aligned} I(Y^X, X) &= H(X) + H(Y^X) - H(X, Y^X) \\ &= H(X) - H(X|Y^X) \end{aligned} \quad (5)$$

which represents the information we can gain about X from Y^X .

Suppose the traffic observed at y is proportional to x at any sampling period. If $Y^X = j$, we can conclude that X equals to a fixed value i . In this case, we have $P(X = i|Y^X = j) = 1$. This, according to Eq. (3), makes the conditional entropy $H(X|Y^X) = 0$. From Eq. (5), we have $I(Y^X, X) = H(X)$, implying that we gain the complete information about X , given Y^X known.

On the contrary, if Y^X is independent from X , the conditional entropy $H(X|Y^X)$ is maximized to $H(X)$. According to Eq. (5), we have $I(Y^X, X) = 0$, i.e., we gain no information about X from Y^X .

In reality, since Y^X records the number of a subset of packets destined to node x , it can not be totally independent of random variable X . Therefore, the mutual information should be valued between the two extremes discussed above, i.e., $0 < I(Y^X, X) < H(X)$. This means that node y can still obtain partial information of X 's traffic pattern. However, a good routing solution should minimize such mutual information as much as possible for any potential observing node. More formally, we should minimize

$$\max_{Y \in \mathcal{V}-X} I(Y^X, X) \quad (6)$$

the maximum mutual information that any node can obtain about X .

3 Penalty-based Routing Algorithm

We propose a penalty-based routing algorithm to achieve our goal of hiding traffic pattern by exploiting the richness of available paths between two nodes in WMN. Specifically, we choose to adopt the *source routing* scheme. Such a choice is enabled by the fact that one node can easily acquire the topology of the WMN it belongs to, which is mid-sized (within 100 nodes) and static. The algorithm operates in three phases, *path pool generation*, *candidate path selection* and *individual packet routing*.

The path pool is generated by repeatedly running shortest path algorithm based on penalties of edges. Every time a node appears in a new path, its penalty value (*tag*) is increased, resulting in lowered chance of being reused in afterwards path generation. This algorithm is designed to balance the needs of routing performance (finding paths with smallest hop count) and preserving traffic pattern privacy (finding disjoint paths). A penalty function serves as the tuning knob to maneuver the algorithm between these two contradictory goals. The penalty for edge (u, v) is defined on penalty values of its vertices: $Penalty(u, v) = \alpha[pow(\gamma, v.tag)] + \beta u.tag$. The parameters (α, β, γ) control how fast the penalty function grows. This in turn, determines how the generated path shifts between “smallest hop-count path” and “disjoint path”. After the path pool is generated, some paths are selected randomly from it to form a combination of diversified routing paths for every destination. For each packet, one routing path is chosen from the selected candidate paths. After routing several packets, this combination is renewed by calling the second phase again. All packets are assigned a randomly chosen path, and all these candidate paths are designed to be disjoint. Hence, the chance that packets are routed in similar paths is small.

Penalty-based routing has been used in existing literature (e.g., [1]), our approach differs from them in applying penalty routing to get better path diversity. Our routing algorithm is also different from the privacy preserving routing in [2], as we mainly consider the trade-off between traffic pattern concealment and routing efficiency, while [2] addresses hiding relaying node identity.

4 Experimental Results

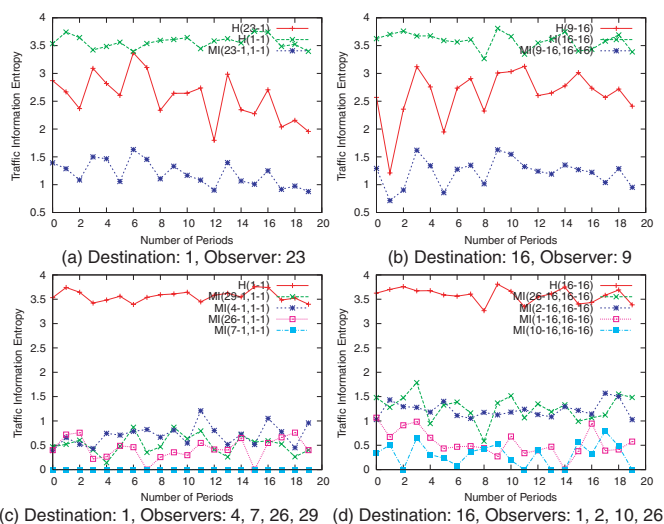


Figure 3. Traffic entropy along time ($\gamma = 1.85$)

We base our simulation on a randomly generated topology ($600 \times 600m^2$) consisting of 30 nodes with transmission range as $250m$. The simulation duration is 400,000 sec. A packet is generated at gateway node 0 each second and its

destination is randomly decided to be one of the other 29 nodes. The simulation time are divided into 20 periods or 1000 intervals. Within each interval, for each destination node x , we count the number of packets that all other nodes y has relayed for x . Then for each period, we independently calculate the traffic entropies $H(X)$, $H(Y^X)$, and mutual information $I(Y^X, X)$ based on their definitions in Sec. 2.

Due to space limit, we only show part of our results. Among all nodes in the network, we choose two sets of nodes. Nodes in the first set $\{1, 6, 11, 15, 23, 24, 25, 29\}$ are close to (2 to 3 hops) the gateway node 0. Nodes in the second set $\{2, 3, 7, 16, 17, 28\}$ are at the edge of the network, 4 to 5 hops away from the gateway. We choose two representative nodes, 1 and 16, out of each set. Fig. 3 shows the variance of traffic entropy and mutual information along the time. In Fig. 3 (a), $H(1-1)$ and $H(23-1)$ denotes the traffic entropy of node 1, and that observed by node 23, respectively. $MI(23-1, 1-1)$ denotes the mutual information node 23 shares with node 1. The same notation rules apply for Fig. 3 (b), where node 16 is the destination, and 9 is the observer. In both pictures, the observing node only shares 40% or less of information about the observed destination node at any sampling period. This observation is further confirmed in Fig. 3 (c)–(d), where we plot the time-variant mutual information that destinations 1 and 16 share with other randomly-chosen observing nodes. These results show that with our algorithm, the destination node is able to consistently limit the proportion of mutual information it shares with the observing nodes.

References

- [1] B. Awerbuch, D. Holmer, C. Nita-Rotaru, and H. Rubens. An on-demand secure routing protocol resilient to byzantine failures. In *ACM Workshop on Wireless Security*, 2002.
- [2] S. Capkun, J. Hubaux, and M. Jakobsson. Secure and privacy-preserving communication in hybrid ad hoc networks. Technical Report IC/2004/104, EPFL-DI-ICA, 2004.
- [3] D. Goldschlag, M. Reed, and P. Syverson. Onion routing for anonymous and private internet connections. *Communications of the ACM*, 42(2):39–41, 1999.
- [4] R. Karrer, A. Sabharwal, and E. Knightly. Enabling large-scale wireless broadband: The case for taps. In *HotNets*, 2003.
- [5] M. G. Reed, P. F. Syverson, and D. Goldschlag. Anonymous connections and onion routing. *IEEE Journal on Selected Areas in Communications*, 16(4):482–494, 1998.
- [6] A. Serjantov and G. Danezis. Towards an information theoretic metric for anonymity. In *Privacy Enhancing Technologies, LNCS*, 2002.
- [7] W. Stallings. *Cryptography and Network Security*. Prentice Hall, 2003.
- [8] X. Wu and B. Bhargava. Ao2p: Ad hoc on-demand position-based private routing protocol. *IEEE Transactions on Mobile Computing*, 4(4):335–348, 2005.