A Framework for Location Privacy in Wireless Networks

Yih-Chun Hu
UC Berkeley
yihchun@cs.cmu.edu

Helen J. Wang
Microsoft Research
helenw@microsoft.com

Abstract

Though an increasing number of wireless hotspots and mesh networks are being deployed, the problem of *location privacy* has been ignored. When a user's location privacy is compromised, an attacker can determine where the user is, and use this information, for example, to stalk or blackmail the user. In existing systems, a user's location can be easily inferred from the signal strengths of packets transmitted from her fixed address. Even if an attacker cannot decode packet contents and addresses, he can correlate different transmissions using a model of the user's movement. In this paper, we argue that location privacy must be a first-class citizen in the design of a wireless communications system. We build a *transaction*-based wireless communication system in which transactions (a single request-response exchange between two nodes) are *unlinkable*; that is, they cannot be correlated. We find that it is even possible to support real-time session-based services such as Voice-over-IP on top of transaction primitives, though with weaker privacy properties. We also identify a number of challenges in providing location privacy in the areas of routing, incentives for multi-hop forwarding, and user- and application-driven tuning of the privacy-performance tradeoff.

Categories and Subject Descriptors: C.0 [**Computer-Communications Networks**]: Security and protection

General Terms: Security, Performance

Keywords: Location privacy, anonymity, wireless networks, ad hoc network routing, security

1 Introduction

In the past few years, we have witnessed the success of numerous wireless communications technologies. Wireless networks now permeate our lives in both time and space: people can compute and communicate almost anywhere, at almost any time. This trend is continuing with the explosive growth of "hotspot" deployment and the new development of multi-hop community networks [6, 1, 25]. Together with improved productivity and greater convenience, these technolo-gies bring with them the rising threat of privacy violations: when a user uses a fixed address, that user can be tracked through the network. In fact, numerous localization techniques [4, 16] have been designed to accurately track a user's movements. Though not all user movements and communications are privacy-sensitive, we must grant users the choice of protecting their privacy when necessary; for example, when a whistleblower leaks information to a reporter or a government agent moves to a secure location in preparation for a possible terrorist attack. Even the network operator cannot be trusted to help keep location information private, since subpoenas, employee theft, or threats of physical violence can be used to extract all available information from the network operator.

We believe that secure, private, unlinkable communications and location privacy represent important problems in wireless networks. Though it may be tempting to try to patch existing protocols to provide some level of privacy, the problems of anonymity and location privacy present fundamental problems that require a precise specification of required privacy properties and purposeful design of a routing architecture to meet those requirements. For example, even if an anonymous routing protocol such as ANODR [15] is used, an attacker can track a user's location through each connection, and associate multiple connections with the same user. When the user arrives at home, she will have left a trail of "packet crumbs" which can be used to determine her identity. In this paper, we explore some of the possible requirements and designs, and present a toolbox of several techniques that can be used to achieve the required level of privacy protection.

In this paper, we show how location privacy can be a first-class design consideration in a wireless communication system. In particular, we argue that privacy on a *transaction* granularity is a fundamental building block of any wireless network. A user could then use this transaction-based privacy support during privacy-sensitive movement or communications. We find that we can support even real-time session-based services such as Voice-over-IP on top of transaction primitives, with some cost in privacy properties.

The rest of this paper is organized as follows. Section 2 describes how an attacker might use localization to track a user. Section 3 provides an overview of our system, which is based on a *transaction*, in which a source sends a stream of data to a destination, and the destination replies with a stream of data. Individual transactions are *unlinkability*; that is, different transactions cannot be associated. It also discusses the use of session based services built on top of transaction primitives. Section 4 describes our design for finding anonymous nodes, addressing and rendezvous, and Section 5 describes challenges specific to multi-hop wireless networks. We describe performance and privacy tradeoffs in Section 6 and related work in Section 7. We then outline some remaining challenges in Section 8 and conclude in Section 9.

(a) Random Waypoint after 100 seconds   (b) Random Waypoint after 200 seconds   (c) Random Waypoint after 500 seconds
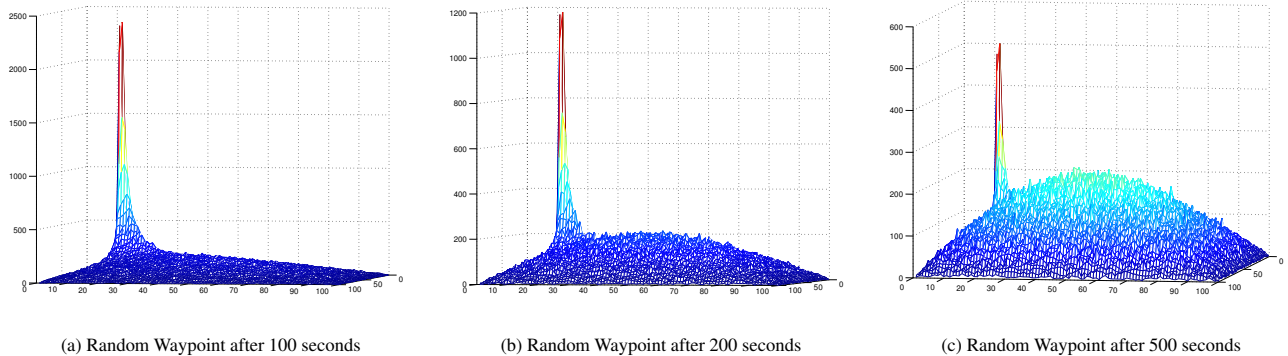
Figure 1: Random Waypoint distribution for nodes starting around (15,5) in a 1000 × 1000 grid. Each figure represents the movement of one million sample nodes, and is plotted with a different z-axis for clarity. After 100, 200, and 500 seconds of mobility, the location of each node was quantized to a 10 × 10 square, and the distribution of nodes is shown as a histogram.

2 User Tracking

A user's several transmissions can easily be correlated if all the transmissions come from the same source address, and the attacker can read the source addresses from each packet. In this section, we demonstrate another source of information by assuming that an attacker cannot decode a user's packets, but still wants to be able to track users around the network. For example, a user may use strong encryption, or the attacker may have insufficient sophistication to build custom hardware for packet decoding, but can measure the signal and noise levels at a variety of locations. Existing localization algorithms (e.g. [4, 16]) allow each transmission to be narrowed down to a specific location. These individual transmissions can then be correlated using a mobility model, as we describe below.

2.1 Attacker Model

The resources available to an attacker wishing to track a user can range from just one machine sitting in a static location in the network to an omnipresent attacker capable of hearing all network transmissions and learning all state kept at a base station. For example, a stalker might only have one static machine and may only be able to infer user location from routing messages and latency to reach the victim's device. On the other hand, a government agency may set up listening posts that are within wireless transmission range of each point in an entire network, and may be able to search the records of the network operator. In this paper, we focus on techniques for defending against the strongest attackers, and we show how some services can only be implemented with a weaker attacker model.

2.2 Localization for User Tracking

When an anonymous user wishes to communicate, she must use a fixed address for some period of time. Whenever the user sends a packet, including acknowledgment packets, an attacker can measure the signal strength of the packet and use an existing localization algorithm to infer the location of the user. In this section, we describe how an attacker might use location information to link different transmissions, and possibly identify the user.

When an attacker accumulates location information for all packets sent in the network, the time at which each packet was sent, and the address from which the packet was sent, the attacker can attempt to correlate different addresses with the same user. For example, if a user was moving along a road at some speed, then a packet further along the same road is more likely sent by that user than is a packet sent from a point that the user recently passed. In particular, the attacker can build a statistical model in which each (location, velocity) pair

corresponds to a probability distribution function of locations at any particular point in the future.

More generally, even when a user cannot be tracked with high certainty, a list of suspects can be rounded up. In particular, when a packet is sent from some location, each user u has some probability p_u of being the sender of the packet. A Maximum Likelihood Estimator can choose the user $\overline{u_{p,\ell}}$ that is most likely to have sent that packet. Even if p_u is small, however, an attacker can create a list of suspects $u_1, u_2, \ldots, u_m$, where $p_{u_1} \geq p_{u_2} \geq \ldots \geq p_{u_m}$ and $\forall u' \notin \{u_1, \ldots, u_m\}, p_{u_m} \geq p_{u'}$. In certain situations, such as for criminal prosecution, a user u_i for sufficiently small i may be subject to search or possibly torture. In such cases, a user may be interested in reducing her p_u to be approximately equal to that of any other node; that is, such that $P[u \in \{u_1, \ldots, u_m\}] \leq \frac{m}{n} + \varepsilon$, where n is the total number of users in the network and ε a small, user-tunable parameter.

2.3 Using Mobility Models

In this section, we demonstrate that an attacker with relatively few location and velocity measurements can still track each user fairly accurately. We analyzed two different types of movement: synthetic movement using the random waypoint mobility model, and actual movement using data from the Seattle bus system. With the synthetic random waypoint mobility model, we generated traces showing how nodes move from waypoint to waypoint. Our goal was to find how long it took for nodes starting at a given point to reach their steady-state distribution; after that time, there is no statistical correlation between the node location and starting point, so an attacker without any intermediate samples has no information. We divided the 1500 m × 300 m and 1000 m × 1000 m areas into 10 m × 10 m sections. For each section, we simulated one million nodes starting in that section, and sampled the locations of those nodes one time per second for 900 seconds.

A representative run of this analysis is presented in Figure 1. After 200 seconds, most nodes fall into the steady-state density distribution; however, even after 500 seconds, a few nodes (around 0.5%) remain very close to their starting location, due to the well-documented problems with the random waypoint model [5, 27, 21], wherein users will sometimes choose very low speeds, causing them to stay stationary for the rest of the simulation.

Our analysis of random waypoint fails to capture three important details of real mobility patterns. First, because we used a synthetic movement pattern, there are artifacts in our analysis peculiar to the synthetic pattern. Secondly, we did not have a model of nodes leaving the network after a period of time. Finally, we assumed we knew only the node's starting location, but not its velocity vector. How-
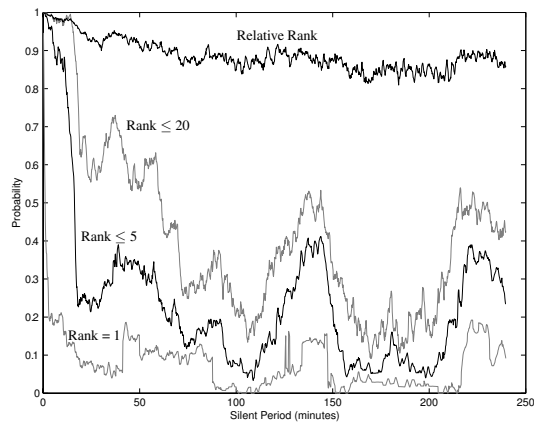
Figure 2: Results of Seattle bus data analysis for one section.

ever, an attacker could learn a transmitter's velocity, for example by using Doppler shifts and triangulation, or by observing several tightly-spaced transmissions. To more accurately capture information available to an actual attacker, we add discretized speed and heading to our probability models. For example, if we are interested in the motion patterns that result in being in the section S defined by $x \in [10, 20), y \in [60, 70)$, then we examine each random movement trace for times at which the node is in that section. In particular, for each discrete timestep t that the node is in the section, we consider the position, speed, and course for every discrete timestep prior to t. For each such prior timestamp t', that node had some position and velocity (x, y, s, θ). We discretize those values, for example to $(\lfloor \frac{x}{10} \rfloor, \lfloor \frac{y}{10} \rfloor, \lfloor \frac{s}{5} \rfloor, \lfloor \frac{4\theta}{\pi} \rfloor)$, and include the time $t - t'$ between that position and the section of interest S. This gives us a Monte Carlo method of determining $P[S | t - t', \lfloor \frac{x}{10} \rfloor, \lfloor \frac{y}{10} \rfloor, \lfloor \frac{s}{5} \rfloor, \lfloor \frac{4\theta}{\pi} \rfloor]$. With this probability distribution, if we know some node is in S and we have location and velocity information for each node at some prior time, we can determine the probability that any node n is the node in S, and thereby find the nodes most likely to have sent a packet from S.

We used this technique to analyze the Seattle bus system mobility data which was derived from the BusView system [11]. We had just over a month of data, which we divided into a training set and a half-day test set[1]. We quantized the time in our data to 10 second intervals, divided the area into square sections of 2000 ft, speed into bins aligned on 5 mph boundaries, and direction into 8 headings. We chose one section at random, and used the training set to examine prior tracks of traffic entering that section, and derived the probability model for $P[S | t - t', \lfloor \frac{x}{2000} \rfloor, \lfloor \frac{y}{2000} \rfloor, \lfloor \frac{s}{5} \rfloor, \lfloor \frac{4\theta}{\pi} \rfloor]$. We then examined the test set and found 2366 quantized time steps t'_i at which a single user was in that section. Based on our probability model, for each user u at each time step t, we computed the probability that u would be in the section at each time step in $\{t'_i\}$. For each (t, t'_i) pair, we sorted all users according to their computed probabilities, and extracted the rank of the user which was actually in the section at that time. We grouped these probabilities according to $t - t'_i$, which represents the amount of time between the (position, velocity) measurement and the observed transmission at S. Within each bin, we plot the fraction of cases where the node's rank was 1 (most probable), ≤ 5 (one of the five most probable), and ≤ 20 (one of the twenty most probable). We also plot the average *relative rank* of the user which was actually in the section at that time, which we define as one minus the ratio between the user's rank and the total number of nodes (for example, if a node ranks 8 of 400, the relative rank is 0.95).

Figure 2 shows the results of this analysis for a single section. (In a real deployment, the same analysis would be made for each section, and we would use the results of the most privacy-compromising sections to determine how long a user is trackable after it transmits). If a user is only worried about being chosen as the highest-probability user, then it will be disassociated with prior movements after around 1 minute for this section and mobility pattern. However, when between 5 and 20 suspects are considered, its prior movements remain correlated for more than 10 minutes. Even after 4 hours, average relative rank shows that considerable information is retained regarding the actual user. Somewhat unexpectedly, in this mobility pattern, the probabilities for this data set are periodic, rather than monotonically decreasing. One possible explanation for this phenomenon is that a single bus tends to service a route for the entire day. When the bus reaches the end of its route, it often pauses and reverses its route. This causes a bus to return near an area at periodic intervals.

In the random waypoint model, after 200 seconds, most nodes can no longer be correlated with their starting position. In the bus mobility pattern, significant location information is retained for periods in excess of 4 hours; however, after 1, 16, and 61 minutes respectively, a node has no higher than a 50% chance of being one of the 1, 5, and 20 most probable nodes.

3 Transaction-Based Wireless Communications

In this section, we give an overview of transaction-based privacy support for wireless networks which fundamentally addresses the location privacy problems raised in the previous section. For clarity, we focus on wireless networks in which packets are routed through well-connected base stations. We address design differences specific to multi-hop wireless community networks in Section 5.

Just as in the existing wireless systems, network providers in our system carry out the tasks of user registration, access control, and billing. While network providers are trusted for these tasks, they are not trusted for any anonymity protection. Therefore, user traffic must appear unlinkable to network providers as well. We detail our **registration and access control schemes** in Sections 4.1 and 4.2.

When nodes can register with a base station through other network clients (such as in a community ad hoc network), a number of additional challenges arise, such as discovering base stations, secure routing, and providing incentives to anonymous, untraceable nodes. We discuss this scenario in more detail in Section 5.

In order to quantify the goals of our system, we introduce the concept of an *unlinkable unit of communications*. We allow an attacker to link any communications within that unit, but we try to prevent the attacker from linking one unit to another unit. These units limit the attacker's ability to track a user. In our system, we provide unlinkability on a *transaction* granularity. A transaction consists of a stream of packets sent from the source to the destination, and another stream of packets sent from the destination back to the source, and is commonly used in applications such as web browsing, database queries, retrieval and sending of email, and DNS queries. Session-based services such as telnet or Voice-over-IP can be considered a series of transactions.

In the previous section, we show that there are two sources of linkability in existing wireless systems: the use of a fixed address for a prolonged period of time and the timing of packet transmissions together with a model of user mobility.

To address the **problem of fixed addresses**, each node in our system frequently changes its IP and MAC addresses. However, as long as a node can be referred to using a single name (such as a domain name), linkability is still retained, since an attacker can repeatedly resolve the domain name to an IP address. This gives the attacker a complete history of the user's IP addresses. We provide a more secure service resolution protocol using an *anonymous bulletin board*. In our service resolution protocol, a user anonymously requests call-

---

[1] In an actual network, an attacker may not have accurate training data; however, we design to defend against a strong attacker which might have representative training data.

backs from the destination. The destination periodically checks this bulletin board and corresponds with the initiator directly. We detail this anonymous rendezvous design in Section 4.4.

To address the **problems of user location inference**, we introduce a *random silent period* for each node, during which that node does not forward or transmit any packets. As we showed in the previous section, location information degrades with time; after a sufficiently long time, two locations can no longer be correlated. The mobility pattern, user density, and user requirements determine a minimum silent period. However, we cannot simply use that silent period between each pair of transactions or the attacker will be able to link the two transactions. To guard against this attack, we use random silent periods.

The use of a silent period necessarily impacts real-time services such as Voice-over-IP. In such services, the latency cost of using mixes (e.g. [7, 23]) would also be prohibitive. As a result, the user must choose between an extremely low quality of service and reduced privacy assurances. In the Voice-over-IP application, for example, a high-privacy telephone call would requires latencies that prohibit natural interaction; rather, a conversation would seem like a series of messages left on answering machines in a bout of phone tag. Other applications, such as an interactive shell, can easily tolerate silent periods, as long as the user does not wish to run any commands during those periods. We explore tradeoffs between privacy and performance in Section 6.

Regardless of which privacy level is chosen by the user, when one user changes addresses, the other user must be updated as to the new address. One way to do this is to include the next address at the beginning of each transaction; for example, if Alice calls Bob, she will update Bob with her next address in her current request, and Bob will respond with his next address in his current reply. Even when Alice and Bob both change addresses at the same time, they will both have the next address from their previous exchange, and therefore will be able to associate their connections. An attacker can use this protocol to link different user addresses. To mitigate this risk, a user might only use encrypted communications with trusted nodes, or may use mixes to hide her address.

Another problem for real-time session-based services is the use of reliable in-order delivery of packets within a transaction. This problem is not fundamental; a simple and inefficient solution is to make each UDP packet a single transaction. This transaction could be specially marked to not require retransmission in case of packet loss.

4 System Design

4.1 Registration

In order to route traffic to a node, the network needs to find which base station that node is associated with. Cellular networks solve this problem with a combination of registration and paging. The same combination can be used in hotspots and community wireless networks, except that periodic registrations present a possible risk to anonymity. On the other hand, without registration, the network cannot guess where a node is, and network-wide paging is neither scalable nor efficient.

In our scheme, we require a user to register in order to send or receive communications. This has the limitation that an anonymous user who is not registered with the network cannot be the destination of a transaction. However, users can use anonymous rendezvous (Section 4.4) indicate the desire to communicate with a node which has not yet registered. When an anonymous user wishes to communicate, or when it changes addresses, it discovers the nearest base station and chooses a cryptographic IPv6 address based on both the base station and the current private key (addressing is further explored in Section 4.3). The user registers with the base station, disclosing the signed certificates (user certificates are discussed in greater detail in Section 4.2). Base stations can be selected using several techniques,

including beacons, flooded beacons, geographical information, and Route Discovery. The details of using Route Discovery to find base stations are further outlined in Section 5.

4.2 Access Control at the Base Stations

Since the deployment of base stations has an associated cost, the network operator must be compensated for the use of the network. One way of achieving this result is to require each authorized network node to carry a piece of secure hardware with a symmetric key shared by all network users. A network of reasonable size cannot completely rely on this mechanism to keep a determined attacker from participating in the network. As a result, we explore techniques for verifying that a user is part of the network.

We use blind signatures in a way that parallels their use in electronic cash: a legitimate user generates a number of public keys, each of which we call an *identity*. The user "blinds" them [8, 9] (so the network provider can sign them without seeing them), and presents them to the network provider. The network provider signs these keys and returns them to the user, who "unblinds" them. Each key signed in this way is a *certificate* which proves that the identity corresponds to a legitimate user, though the particular user cannot be determined.[2]

A problem with this approach is that any legitimate user can act as a certificate signing service simply by passing requests to the network provider. If the service provider is paid for each certificate, or if the number of certificates that the service provider will sign for any client is reasonably limited, this may not be a problem. Otherwise, if the network is paid for on a flat rate basis, we divide time into time slots, and require the user to include the time slot number in each public key for which the user requests a signature. Before the beginning of each month, the user presents one identity for each time slot in that month, and the network operator signs each identity with a different public/private key pair (one key pair for each time slot). For example, if the time slots are 5 minutes each, the user will present around 4500 certificates each month, and the network operator will sign each certificate with a different key. A user therefore cannot allow two different devices to connect at the same time. This method of metering is analogous to the exchange of a physical token which provides access; one cell phone subscriber can lend the phone to different people at different times, but only one user can use the network at a time.

Depending on the network operator's revenue model, different approaches certificate generation can be used; however, we caution against allowing users to return unused certificates for a credit. Otherwise, the network operator can compile a list of users who were active during each time slot, which could seriously jeopardize the anonymity of any user which did register during a time slot.

4.3 Addressing

In our addressing design, we aim to avoid collisions in the address space while still retaining anonymity. Our addresses are statistically unique cryptographically verifiable IPv6 addresses [20]. These addresses contains three components: the network prefix, base station ID, and a hash of the user's identity (public key). For example, we might use the 10-bit site-local prefix[3], a 10-bit base station ID, and a 108-bit hash of the user's identity. Though a hash collision (two identities with the same hash) on a 108 bit hash can be found with just 2^{54} work, an node creating such a collision must generate both identities; for any other node to generate such a collision requires 2^{107}

---

[2]Using RSA, if the network operator has public key e, private key d, and public modulus n, and the user wants to have m signed by the operator, she chooses r at random between 1 and n, and computes $b = mr^e \mod n$. The network operator signs $b^d = (mr^e)^d \mod n = rm^d \mod n$. The user divides out her random value r to get the certificate $b^d/r = m^d \mod n$.

[3]A globally-unique prefix could be used if assigned. IANA intends to assign 32-bit and 48-bit prefixes [2] (for networks and organizations respectively), which, when combined with a 10-bit base station ID, provides 86 and 70 bits of security respectively.

work on average, which is computationally infeasible. By encoding base station ID into addresses, we can support mobility without using an entity such as a Mobile Location Register, which keep track of the current location of mobile nodes in cellular networks. When a node moves from one base station to the next, the node uses original base station as a Mobile IP Home Agent [14], and registers a new geographical address with that base station.

One risk of having a geographical address is that a node is trivially locatable to an attacker with few resources. If, based on the application, this risk is considered too great, the network operator can allow the use of *unassociated* address. With such an address, the geographical prefix is set to a value not associated with any base station. The network operator then provides a centralized Mobile Location Register for these addresses (as in done in cellular networks) or serves as a centralized IPv6 Home Agent for these addresses. We defer the details of these schemes to prior work. A disadvantage to both of these approaches is that they require the network operator to keep information about where each node is; if such information is logged for a long period of time, it could be used to compromise node privacy.

4.4 Anonymous Rendezvous

We employ an anonymous bulletin board as a means of rendezvous to allow a pair of secretive communicators to find the current address of the other communicator. Before each round of communications, the initiator needs to use the bulletin board to indicate its address and desire to communicate to a particular destination node, The initiator does this by posting a *callback request* on the bulletin board. Then, when the destination node sees the request on the bulletin board, it communicates with the initiator using its address directly. (We assume the bulletin board can be trusted to not modify, add, or remove any postings in an unauthorized way.) To achieve unlinkability across several rounds of communications, each initiator generates a token chain for each destination node, and sends the token chain to the destination node through a secure (out-of-band) channel before any rounds of communications take place. The ith token is used by a request for ith attempt to communicate. In other words, these tokens are used consecutively, and one is used per rendezvous attempt. Each token is a triple of (round nonce, encryption key, authentication key) where the round nonce is a random number generated for uniquely identifying a round, the encryption key is used to encrypt the initiator's address in the request, and the authentication key is used to verify the authenticity of the address[4]. We denote the ith triple as (η_i, K_i, K_i'). The callback request for round i takes the form $(\eta_i, E_{K_i}(\text{addr}), \text{HMAC}_{K_i'}(\text{addr}))$.

Each node periodically requests the list of all callback requests (or, alternatively, the list is periodically flooded), and searches the callback table for round nonces (η_i) which it shares with nodes with which it communicates. If it finds such a request, it decrypts the address with the encryption key, checks the authenticity with the authenticator, and, if it is authentic, begins communicating with the address specified in the callback request.

The initiator may enter its silent period before it receives a response to its callback request. We therefore need a way to remove a callback request. Since the entire callback request list is public, however, a malicious user can reinsert removed callbacks. We now describe how callbacks can be added and removed securely. When generating the callback request, each node also generates a *revocation value* r. The callback request for round i takes the form

$$(\eta_i, E_{K_i}(\text{addr}), \text{HMAC}_{K_i'}(\text{addr} \mid\mid H(H(r))), H(H(r)))$$

where H is a one-way hash function. In order to post this callback request, the node needs $H(r)$, and to revoke this callback request,

the node needs r. A malicious user cannot revoke a callback request that it did not generate, because it cannot invert the one-way hash function. In addition, once a callback request is revoked, an attacker cannot repost it verbatim, because it does not know the correct $H(r)$ value (assuming that communications between each node and the bulletin board are encrypted). Finally, if the attacker chooses a different revocation value r', the bulletin board will post the request, but the destination will find that the authenticator is invalid.

5 Multi-Hop Routing

In some wireless networks, a client can join the network even when it is not in direct wireless transmission range of a base station. Such networks, commonly called *ad hoc networks*, rely on other clients to forward packets to enable communications between nodes not directly in wireless transmission range of each other.

5.1 Base Station Discovery

The most straightforward Base Station Discovery requires each base station to perform a periodic hop-limited flood to allow users to decide which base station to register with.

5.2 Secure Routing

In order to prevent a malicious node from disrupting routing, the ad hoc networking research community continues to explore secure routing protocols [24, 28, 3, 12, 22]. Any of these protocols could be slightly modified to work within our network architecture. Depending on user requirements, an anonymous routing protocol, such as ANODR [15], may provide more desirable properties.

To show how we might adapt a secure routing protocol to our architecture, we first consider an on-demand routing protocol. We must allow a base station to reply to any Route Discovery; however, because there are a limited number of base stations and their keys are easily pre-distributed, they can easily send and authenticate RREPs. We must also adapt the existing protocols to choose routes that can traverse a base station; in hop-by-hop routing, one technique is to have a node forward the packet to the nearer of the destination node or the nearest base station, unless the packet is sent by a base station, in which case we never route it to a base station. When a base station receives a packet, it forwards it to the appropriate base station based on the destination's geographical address. The base station with which the destination is associated then encapsulates the packet so that forwarding nodes know that the packet has already traversed the base station. Periodic protocols can use the same forwarding logic.

5.3 Incentives for Anonymous Routing

Many researchers have observed that nodes in an ad hoc network do not always have an incentive to forward packets for other users. As a result, several schemes have been proposed to incentivize forwarding or punish misbehavior (e.g. [29, 19, 13]). However, in an environment where users are completely anonymous and can quickly change between identities, such systems can quickly break down. A punishment mechanism would not work, since any time a user was punished, she could simply change identities. When forwarding incentives are provided, payment must be made either before the packet is forwarded or after the packet is forwarded. In a prepayment scheme, a node can take the payment and fail to forward the packet. In a postpayment scheme, a node can request forwarding and then fail to pay. Without the involvement of a trusted authority that can act as an escrow and verify packet delivery, payment schemes fail. One solution to this problem is to use the base station to escrow payment. We require that the base station learn which nodes are forwarding packets, for example through a link-state routing table, a source route, or information piggybacked on Route Discovery.

---

[4] All tokens can be generated from a single master key $\mathcal{K}$ using a pseudorandom function $\mathcal{F}$: $\eta_i = \mathcal{F}_{\mathcal{K}}(3i)$, $K_i = \mathcal{F}_{\mathcal{K}}(3i+1)$, and $K_i' = \mathcal{F}_{\mathcal{K}}(3i+2)$.

A packet sent from one mobile host to another first traverses a number of nodes to reach the source's base station, travels to the destination's base station, and traverses a number of nodes to reach the destination's base station. When the packet moves from the source to the source's base station, each forwarding node can be credited when the unique, authenticated packet arrives at the base station. When the packet travels from the destination to the destination's base station, the destination base station can credit forwarding nodes when it receives an acknowledgment from the destination. In this case, a forwarder at most can be underpaid by one window size, and a forwarder can easily limit the window size by dropping packets until an acknowledgment comes back. Payments for forwarding can be made using electronic cash. These payments can be posted to a globally accessible site, with each payment made accessible to the private key of the forwarding node. Since some forms of anonymous electronic cash require a three-way exchange [10], we piggyback the required exchange on the data and acknowledgment packets sent as part of the transaction. When the transaction is ended or times out, the base station can either send the payments or post them publicly.

6 Performance Tradeoffs

6.1 Anonymity-Efficiency Tradeoffs

Because of the inefficiency in providing anonymity and location privacy for all packets, a protocol providing anonymity may offer a user the option of operating in "normal" mode. In any protocol where anonymity is an optional service (provided for some packets and not others), there is a difficulty in establishing an intuitive user interface that allows a user to understand the privacy implications of their actions. One option is to allow a user to switch her device between privacy mode and normal mode [17]; in privacy mode, the device conceals each transaction with a long randomized delay. When the device is switched to normal mode, it registers with the network after another randomized delay. The difficulty with this approach is that it takes too conservative an approach towards parallel transactions. Though a user may wish to avoid correlation between two transactions (for example a money transfer and an anonymous web browsing session), many anonymous sessions should be parallelizable. Another option is to allow a user to create a number of pseudo-identities (other than the user herself). In this model, any actions performed under any single pseudo-identity can be parallelized, and the user's device would schedule transactions in order to ensure that different pseudo-identities could not be correlated.

In certain areas, a user's location will automatically disclose her identity. For example, when a user is in her home or office, any communications emanating from these areas will be interpreted as hers with high probability. In addition, the user may wish to set up "restricted" areas (in which she does not participate in anonymous communications) around these locations, since an attacker could potentially track the motion of the user towards or away from her home or office, and infer which user sent the communication.

6.2 Mixes for Higher Security

Though our system can provide location privacy and unlinkable communications, it is still vulnerable to traffic analysis, where an attacker gathers information about pairs of nodes which communicate. In wired networks, researchers have envisioned *mixes* [7, 23] to remove the correlation between sources and destinations. In our architecture, we can use the base stations as mixes; if sufficient traffic traverses the base stations, the base stations alone can provide strong protection against traffic analysis. However, this mixing is insufficient, since the base stations may not be trusted. In addition, not every application may require mixing, because of the additional latency inherent in a mix. In our architecture, we include a separate application-layer mixing service. An application can choose a chain of mixes through which to conduct its transaction. Prior work shows that unless all mixes are compromised, strong protection against traffic analysis is possible.

7 Related Work

A number of researchers have proposed systems combining cellular infrastructure with ad hoc networks [18, 26]. We build on such architectures, focusing on the properties of anonymity and location privacy.

Security and incentives for participation in ad hoc network routing have also been extensively studied; for example, a number of routing protocols (e.g. [24, 28, 3, 12, 22]) and routing incentives (e.g. [29, 19, 13]) have been developed. In our work, we build on these protocols to provide such services in the face of anonymity and location privacy, and in conjunction with a trusted cellular infrastructure.

Kong and Hong proposed an anonymous routing protocol [15]. This protocol makes use of mix routing [7] to allow anonymous communications (that is, the destination does not know who the source was). However, the protocol does not provide location privacy, since the several transmissions of a node can be correlated. Eventually, the node may transmit from a location which reveals its user's identity. If the user was forwarding packets for a large number of other users, its communications will still remain anonymous.

8 Future Work

In this paper, we have outlined the privacy problems associated with wireless networks, and sketched an architecture that could support anonymous routing and location privacy in such networks. However, many open problems remain in this area:

- **Mobility modeling.** When nodes use shorter silent periods, they can have higher performance and lower latency; however, to accurately determine a safe silent period, we need a better model for mobility. Such a model could be a synthetic model or a large number of actual traces.

- **Formal framework for user tracking.** Though we introduce techniques for tracking user mobility in Section 2, we have not provided a formal model to evaluate the optimality of this tracking. Such a formal model would allow better understanding of the privacy provided by a given choice of silent period.

- **Privacy objectives.** In this paper, we have focused on an attacker with tremendous power— he can hear every packet transmitted from any point in the network, and measure the location from which each of those packets is transmitted. Most users do not need this level of privacy; better model of an attacker and corresponding reasonable privacy expectations could lead to higher-performance protocols.

- **Communications pattern.** In this paper, we focus on transactions as a communications model. Though this model fits many kinds of network usage, some best-effort services such as Voice over IP cannot operate within this model. Development of additional transport protocols which are compatible with the required silent periods, as well as able to support payment for forwarding, would bring more applications to such networks.

- **User interface.** Providing location privacy, uncorrelated traffic streams, and high performance using published techniques seems to require the parallelization of related traffic streams. A user may then be required to dictate policy about which streams can be associated without damaging the user's anonymity. However, since users are often the weakest link in systems security, an intuitive user interface is of paramount concern to the design of a secure system.

- **Secure routing protocols.** Another problem is the design of efficient routing protocols that have both strong security and high network performance. Though security extensions have been

designed for several existing protocols, many of these extensions remove important performance optimizations. Optimistic approaches may provide a better trade-off between security and performance.

- **Forwarding incentives.** Though a number of researchers have proposed methods for incentivizing packet forwarding, they do not consider the case in which the forwarding nodes wish to remain anonymous.

9 Conclusions

In this paper we have introduced the problems of anonymity and location privacy in wireless networks. We considered various attacker strengths and showed how user mobility and density drives protocol design, examining both synthetic and real-world movement traces. We proposed a novel communications abstraction, based on a transaction, and designed a system that provides unlinkable transactions in a network with centralized base stations. We also identified and presented first solutions to a number of challenges in providing location privacy and unlinkability in the areas of routing, incentives for multi-hop forwarding, and user- and application-driven tuning of the privacy-performance tradeoff.

Acknowledgements

References

[1] Daniel Aguayo, John Bicket, Sanjit Biswas, Douglas S. J. De Couto, and Robert Morris. MIT Roofnet Implementation. Technical report, MIT, August 2003. Available at `http://pdos.lcs.mit.edu/roofnet/design/`.

[2] Takashi Arano, Thomas Narten, and David Kessens. IPv6 Address Allocation and Assignment Policy. `http://www.iana.org/ipaddress/ipv6-allocation-policy-26jun02`, June 2002.

[3] Baruch Awerbuch, David Holmer, Cristina Nita-Rotaru, and Herbert Rubens. An On-Demand Secure Routing Protocol Resilient to Byzantine Failures. In *ACM Workshop on Wireless Security (WiSe)*, September 2002.

[4] Paramvir Bahl and Venkata N. Padmanabhan. RADAR: An In-Building RF-Based User Location and Tracking System. In *Proceedings of the Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2000)*, volume 2, pages 775–784, March 2000.

[5] Christian Bettstetter and Christian Wagner. The Spatial Node Distribution of the Random Waypoint Mobility Model. In *Proceedings of the German Workshop on Mobile Ad-Hoc Networks (WMAN)*, number P-11 in GI Lecture Notes in Informatics, pages 41–58, March 2002.

[6] Pravin Bhagwat, Bhaskaran Raman, and Dheeraj Sanghi. Turning 802.11 Inside-Out. In *Proceedings of the Second Workshop on Hot Topics in Networks (HotNets-II)*, November 2003.

[7] David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 4(2), February 1982.

[8] David Chaum. Blind Signatures for Untraceable Payments. In *Advances in Cryptology — CRYPTO '82*, pages 199–203, 1983.

[9] David Chaum. Security Without Identification: Transaction Systems to Make Big Brother Obsolete. *Communications of the ACM*, 28(10):1030–1044, October 1985.

[10] David Chaum, Amos Fiat, and Moni Naor. Untraceable Electronic Cash. In *Advances in Cryptology — CRYPTO '88*, pages 319–327, 1990.

[11] D.J. Dailey, G. Fisher, and S. Maclean. Busview and Transit Watch: an Update on Two Products from the Seattle Smart Trek Model Deployment Initiative. In *Sixth Annual World Congress on Intelligent Transport Systems*, November 1999.

[12] Yih-Chun Hu, Adrian Perrig, and David B. Johnson. Ariadne: A Secure On-Demand Routing Protocol for Ad Hoc Networks. In *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking (MobiCom 2002)*, pages 12–23, September 2002.

[13] Markus Jakobsson, Jean-Pierre Hubaux, and Levente Buttyán. A Micro-Payment Scheme Encouraging Collaboration in Multihop Cellular Networks. In *Proceedings of the Seventh International Financial Cryptography Conference (Financial Cryptography 2003)*, January 2003.

[14] David B. Johnson, Charles E. Perkins, and Jari Arkko. Mobility Support in IPv6. Internet-Draft, draft-ietf-mobileip-ipv6-24.txt, June 2003. Work in progress.

[15] Jiejun Kong and Xiaoyan Hong. ANODR: ANonymous On Demand Routing with Untraceable Routes for Mobile Ad-hoc Networks, June 2003.

[16] Andrew M. Ladd, Kostas E. Bekris, Algis Rudys, Guillaume Marceau, Lydia E. Kavraki, and Dan S. Wallach. Robotics-Based Location Sensing using Wireless Ethernet. In *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking (MobiCom 2002)*, pages 227–238, September 2002.

[17] Scott Lederer, Jennifer Mankoff, and Anind K. Dey. Who Wants to Know What When? Privacy Preference Determinants in Ubiquitous Computing. In *Short Talk in the Extended Abstracts of CHI 2003, ACM Conference on Human Factors in Computing Systems*, pages 724–725, April 2003.

[18] Haiyun Luo, Ramachandran Ramjee, Prasun Sinha, Li Erran Li, and Songwu Lu. UCAN: a unified cellular and ad-hoc network architecture. In *Proceedings of the Ninth Annual International Conference on Mobile Computing and Networking (MobiCom 2003)*, pages 353–367, September 2003.

[19] Sergio Marti, T.J. Giuli, Kevin Lai, and Mary Baker. Mitigating Routing Misbehaviour in Mobile Ad Hoc Networks. In *Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking (MobiCom 2000)*, pages 255–265, August 2000.

[20] G. Montenegro and C. Castelluccia. Statistically Unique and Cryptographically Verifiable (SUCV) Identifiers and Addresses. In *Symposium on Network and Distributed Systems Security (NDSS 2002)*, February 2002.

[21] William Navidi and Tracy Camp. Stationary Distributions for the Random Waypoint Mobility Model. *IEEE Transactions on Mobile Computing*, 3(1):99–108, January–March 2004.

[22] Panagiotis Papadimitratos and Zygmunt J. Haas. Secure Routing for Mobile Ad Hoc Networks. In *SCS Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS 2002)*, January 2002.

[23] Michael K. Reiter and Aviel D. Rubin. Crowds: anonymity for Web transactions. *ACM Transactions on Information and System Security*, 1(1):66–92, 1998.

[24] Kimaya Sanzgiri, Bridget Dahill, Brian Neil Levine, Elizabeth Royer, and Clay Shields. A Secure Routing Protocol for Ad hoc Networks. In *Proceedings of the 10th IEEE International Conference on Network Protocols (ICNP '02)*, November 2002.

[25] Rudi van Drunen, Dirk-Willem van Gulik, Jasper Koolhaas, Huub Schuurmans, and Marten Vijn. Building a Wireless Com-

munity Network in the Netherlands. In *Proceedings of the USENIX 2003 Annual Technical Conference*, pages 219–230, June 2003.

[26] Xiaoxin Wu, G.-H. Gary Chan, and Biswanath Mukherjee. MADF: A NovelApproach to Add an Ad-Hoc Overlay on a Fixed Cellular Infrastructure. In *Proceedings of the IEEE Wireless Communications and Networking Conference 2000 (WCNC 2000)*, September 2000.

[27] Jungkeun Yoon, Mingyan Liu, and Brian Noble. Random Waypoint Considered Harmful. In *Proceedings of the Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2003)*, volume 2, pages 1312–1321, April 2003.

[28] Manel Guerrero Zapata and N. Asokan. Securing Ad Hoc Routing Protocols. In *Proceedings of the ACM Workshop on Wireless Security (WiSe)*, pages 1–10, September 2002.

[29] Sheng Zhong, Jiang Chen, and Yang Richard Yang. Sprite: A Simple, Cheat-Proof, Credit-Based System for Mobile Ad-Hoc Networks. In *Proceedings of the Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2003)*. IEEE, April 2003.