Mobile Networks and Applications 10, 315–325, 2005

Enhancing Location Privacy in Wireless LAN Through Disposable Interface Identifiers: A Quantitative Analysis

MARCO GRUTESER and DIRK GRUNWALD
Department of Computer Science, University of Colorado, Boulder, CO

Abstract. The recent proliferation of wireless local area networks (WLAN) has introduced new *location* privacy risks. An adversary controlling several access points could triangulate a client's position. In addition, interface identifiers uniquely identify each client, allowing tracking of location over time. We enhance location privacy through frequent disposal of a client's interface identifier. While not preventing triangulation per se, it protects against an adversary following a user's movements over time. Design challenges include selecting new interface identifiers, detecting address collisions at the MAC layer, and timing identifier switches to balance network disruptions against privacy protection. Using a modified authentication protocol, network operators can still control access to their network. An analysis of a public WLAN usage trace shows that disposing addresses before reassociation already yields significant privacy improvements.

Keywords: location privacy, wireless LAN

1. Introduction

Wireless Local Area Hotspot Networks (WLAN) based on the IEEE 802.11b standard [19] have been widely deployed throughout university and corporate campuses, airports, conference centers, and even coffee shops. WLAN interfaces are also incorporated into an increasing number of mobile devices including cell phones, PDAs, and digital cameras. Researchers and the automotive industry even expressed interest to build WLAN capabilities into vehicles (e.g., [29]). Applications could range from gas payments, over download of entertainment content, to collision warnings through ad-hoc communication between vehicles [14]. Along with its conveniences such as enabling greater user mobility, this widespread deployment also introduced significant privacy risks that may affect user acceptance of hotspot services.

While most recent work addressed *content* privacy (i.e., eavesdropping) risks [9,36], WLAN networks also introduce greater *location* privacy risks. Every client is uniquely identified by its default MAC address, and the location of a client is at least approximately known through the location of the associated access point. Furthermore, researchers have determined the location of a client with up to 1m accuracy by collecting signal measurements from multiple access points[1] and by applying signal processing algorithms that take into account the propagation characteristics of the environment [3,10,23,33]. While some researchers have proposed systems for controlling the dissemination of location information, the challenge of protecting WLAN users against untrustworthy access point operators has, to our knowledge, not been addressed. Less trustworthy access points may include the access points encountered when roaming in foreign networks and other parties that can overhear wireless communications.

This paper presents short-lived, disposable MAC addresses – an approach that effectively reduces the opportunities for location tracking. Specifically, it

- defines an attack model for location privacy risks inherent in WLAN,
- describes the design and implementation challenges of disposable MAC addresses in WLAN and its integration into subscriber authentication services, and
- quantitatively analyzes the location privacy protection based on WLAN usage traces from a conference setting.

The remainder of this paper is organized as follows. First, Section 2 assesses the location privacy risks inherent in WLAN usage. Section 3 the presents the design challenges introduced by temporary interface identifiers. We evaluate disposable MAC addresses using mobility traces collected from a public wireless network in Section 4 and discuss the results in Section 5.

2. Wireless LAN location privacy challenges

In the United States, privacy risks related to *location* information have been identified in the *Location Privacy Protection Act of 2001* [26]. While public disclosure of location information enables a variety of useful services such as improved emergency assistance, it also exhibits significant potential for misuse. For example, location information can be used to *spam* users with unwanted advertisements or to learn about users medical conditions, alternative lifestyles or unpopular political views. Inferences can be drawn from visits to clinics,

[1] Not all systems implement the location calculation algorithms at the access point, but RADAR [3] demonstrated the feasibility of this approach. Furthermore, the usage of client-centric location sensing mechanisms does not prevent access point operators from implementing their own location sensing systems.

doctors' offices, entertainment districts, or political events. Such conclusions can be particularly annoying for subjects if inaccurate. In extreme cases, public location information can lead to physical harm, for example in stalking[2] or domestic abuse scenarios.

Commercial Hotspot wireless services are typically offered by subscription similar to a mobile phone contract. Thus, a user or data subject enters into a contractual agreement with one service provider who typically provides accounting of service usage and monthly billing. We refer to this service provider as the *home service provider (HSP)*. This provider may enter roaming agreements with other service providers, the *foreign service providers (FSP)*. In this context, the user only has a direct contractual agreement with the HSP and may wish to protect identity information from FSPs and eavesdroppers.

Generally, a location privacy threat describes the risk that an untrusted party can locate a transmitting device *and* identify the subject using the device. Wireless LAN networks pose especially serious location privacy threats for the following reasons:

*Untrusted network operators***.** Establishing a contractual relationship that regulates privacy issues would be cumbersome for spontaneous network access through small providers. Especially, in a "community network" or "lilypad" [28] scenario where the actual network operators might be individuals unknown to the network clients. In addition, third parties can easily eavesdrop, since WLAN access points have become affordable for individuals to acquire and operate.

*High Density of access points***.** Population centers, where most people spent the larger part of their life, already exhibit a high density of WLAN access points [18]. Figure 1 shows the concentration of some of the access points scanned *via* "war driving" in the Chicago area [38]. Thus, the spatial and temporal coverage of the location privacy threat is greatly extended. Note that location tracking of a network client *does not* require association to an access point. If a client sends messages, any access points in radio range (and on the same channel) can overhear these messages for location tracking purposes.

*Precise positioning technology***.** The radio signal properties of a WLAN system allow relatively precise determination of a client's position. When an access point receives a signal from a client, the client position is with high probability within the typical range of an IEEE 802.11b system, about 50–100 meters. An active thread of research [3,10,23] improves upon this accuracy through triangulation based on the signal strength and signal-to-noise ratio received at multiple cooperating access points. These systems can achieve accuracies of up to 1 meter in an indoor environment.

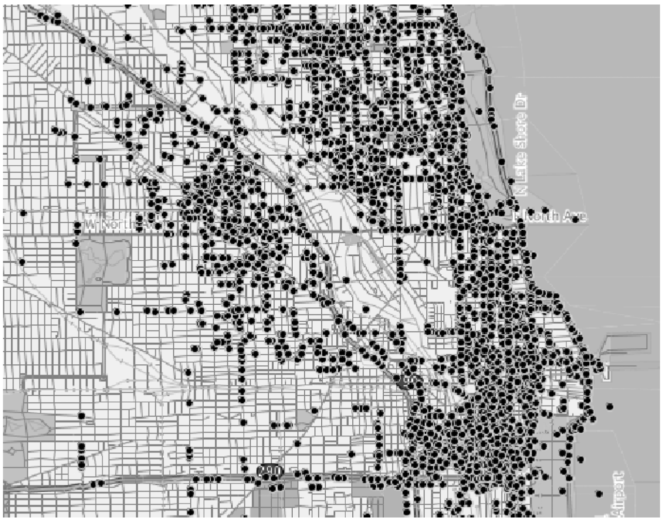

Figure 1. Density of WLAN access points in downtown Chicago.

*Interface identifiers***.** Networks based on the IEEE 802 family of standards use 48-bit interface identifiers (MAC addresses). The most significant 24 bits of an address contain the Organizationally Unique Identifier. This identifier is assigned by the IEEE to hardware manufacturers to ensure uniqueness of interface identifiers. The manufacturer assigns unique values for each network card to the remaining 24 bits and stores the complete interface identifier in the firmware.

By default, the network card driver reads and uses the stored address, thus the identifiers typically persist as long as the same network card is used. In addition, mobile computing devices are rarely shared between different people, thus a particular interface identifier usually correlates with a certain data subject. This assumption enables an adversary to reidentify a subject that has been previously observed; thus, movement patterns can be created. It also allows merging of location data collected from different access points.

Compared to mobile phone systems, WLAN introduces increased location privacy risks, because untrusted parties can easily obtain the equipment to eavesdrop on WLAN clients and remote positioning technology is more precise.[3] We also expect that WLAN interfaces will become much more pervasive in the near future, thus allowing the tracking of our position even when we are not using notebook computers.

While information from many access points must be merged to allow continuous location tracking in larger areas (due to the relatively short range), WLAN still poses a significant threat. It is plausible that access point records are subpoenaed, that an organization controls a large number of

[2] A recent stalking case [11] illustrates the exploitation of positioning technology for personal vendettas. A person mounted a GPS receiver and digital cellular transmitter on his former girlfriend's vehicle to track her whereabouts.

[3] Mobile phone service providers are facing difficulties implementing the E911 requirements through remote mechanisms that do not modify the phone units [30].

access points, or that adversaries with access points in "interesting" locations will offer the observed MAC addresses on a black market similar to the "spotting market" feared by Phil Agre [1].

3. System design

A location-privacy enhancing system that protects clients from untrustworthy access points, must either curb the access points' ability to determine clients' location or prevent access points from identifying clients. We observe that a client has little influence on the location sensing capabilities of an access point – other than ceasing all signaling. Therefore, we concentrate on curbing identification of clients.

The system aims at reducing identification risks through temporary interface identifiers. Permanent identifiers provide key information to the adversary. For example, the adversary might have prior knowledge of the interface identifier for a particular user. In this case, the interface identifier reveals the identity of the user. Even without this knowledge, the identifier enables tracking of a user's movements, because messages sent from different locations at different times carry the same identifier. If at any point the adversary can determine the identity of a user, he can map all past and future movements to this user.

Short-lived interface identifiers mitigate these risks. Even when the adversary can determine the user of a particular identifier, the adversary will likely lose track when the identifier is disposed.

3.1. Goals

*Unlinkable identifiers***.** The primary goal of this system is to enhance location privacy through short-lived identifiers. The privacy protection is significantly reduced if an adversary can determine that an old identifier and a new identifier belong to the same client (i.e., link the old and new identifier). Therefore, the key design challenge is to minimize clues that would aid an adversary in performing this linking attack.

*Minimal network disruption***.** Ideally, switching the interface identifiers should cause no or minimal network disruptions to the client user and other users on the network. Network disruptions include performance penalties such as dropped packets or brief disassociations from the access point.

*Applicability***.** The solution should be readily applicable to current IEEE 802.11b access points and client adapters. We do not consider sophisticated hardware such as directional antennas that might thwart position determination through triangulation. In addition, location privacy should not prevent a legitimate Home Service Provider to authenticate network users for access control and accounting purposes.

Incorporating the concept of short-lived interface identifiers into WLAN creates several key design challenges: identifier selection, identifier uniqueness, and integration with port authentication.

3.2. Address selection

Fundamentally, a newly selected MAC address must be unlinkable to the previous address; that is, an adversary cannot determine from the addresses that they both belong to the same client. Furthermore, the address must be valid under the IEEE 802 standard. Valid client addresses are 48 bits long, start with an OUI, and exclude reserved addresses such as broadcast (`FF:FF:FF:FF:FF:FF`) and multicasts.

We create valid and unlinkable addresses based on a forward chain of MD5 hashes started with an unpredictable random seed [13] (the last hash constitutes the input for the next application of the hash function). Only part of the 128-bit hash is used to generate the address, thus an adversary never receives access to the full hash value. A small number of bits – in our implementation we used the 3 least significant bits – serves as an index into a short table of valid OUIs from common wireless network interface card (NIC) manufacturers. OUI information can be obtained from the IEEE OUI assignment list [20]. The next 24 bits of the hash are concatenated to the OUI to create the full MAC address. The algorithm only chooses OUIs from common wireless cards, so that the randomly selected address appears unsuspicious and indistinguishable from real addresses. Network equipment manufacturers often subdivide their OUI address range into several classes for different network products. The address selection procedure could be further refined to only use address blocks used for wireless NICs. In the remainder of this paper an informal reference to a "randomly chosen address" describes this address selection procedure.

3.3. Link layer duplicate address detection

When several clients randomly choose their addresses, collisions are possible. Using the *birthday paradox* we can calculate an upper bound for the probability of an address collision over a time interval t as

$$p(t) = 1 - \left(1 - \frac{n(n-1)}{2^{b+1}}\right)^{tf}$$

where b is the number of randomly chosen address bits, f is the frequency of address switches, and n is the number of clients on the same LAN.

Figure 2 illustrates that address collisions with 27 randomly chosen bits (assuming 8 different OUIs) are improbable for a small number of clients. However, duplicate addresses become a problem for larger networks that serve hundreds of clients through several connected access points.

Duplicate address detection mechanisms mitigate the problems caused by address collisions. However, standard duplicate address detection systems such as *gratuitous ARP* [35] assume the assignment of unique interface identifiers to enable

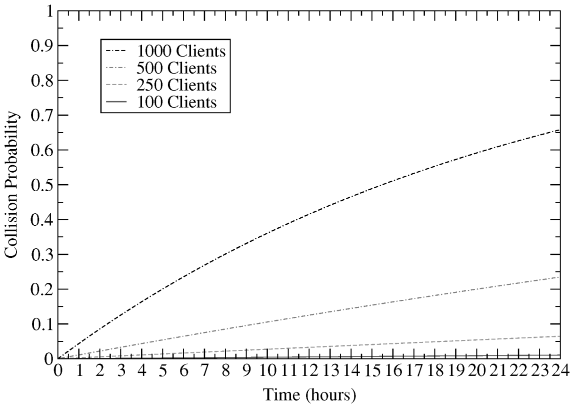

Figure 2. Upper bound for the probability of collision on 27-bit addresses in a given time interval if all clients switch their addresses every 5 minutes.

detection of collisions among network addresses. In comparison, the link layer collision detection mechanisms need to infer duplicated interfaces identifiers without collision-free access to the network. Furthermore, these mechanisms do not take privacy into account.

Since the hidden node problem foils a completely passive approach – promiscuously listening to the network medium – we chose an active ARP-based approach. However, sending a reverse ARP request for the new interface identifier with the current identifier as a source address reveals exactly the information we seek to hide: the link between new and old address. Therefore, we extend this mechanism with a double address switch. First, the client writes a randomly chosen address into the source field of a reverse ARP request. The actual request is for a second randomly chosen address. After sending the packet, the client listens for replies to the first address. A reply indicates an occupied second address; thus, the client repeats the request with a new randomly chosen address until an available address is found.

This protocol still allows address collisions on the first randomly chosen address. However, the client uses this address only to transmit a small number of reverse ARP requests. Therefore, no significant network disruptions should occur.

3.4. Integration with port authentication

Hotspot service providers may need to restrict network access to a known set of clients, say, paying customers. Standards such as IEEE 802.1x, EAP-TLS, CHAP, and RADIUS serve this need. These authentication protocols, however, introduce other identifiers besides MAC addresses that enable eavesdroppers and foreign access points to track users. In EAP-TLS [12], for example, the client (supplicant, in IEEE 802.1x terminology) first transmits a known client identifier and then provides a certificate with the client's public key. Both uniquely identify the client.

The following modifications to such protocols overcome the apparent tension between privacy and access control. The use of one-time, temporary identifiers in the foreign network and over the wireless medium prevents eavesdroppers from tracking users. Splitting the identifier into a network and client part protects the identity of roaming clients[4] from foreign access points, while still allowing less trusted access points to forward authentication requests to the correct network.

Table 1 outlines an EAP-TLS-derived protocol that incorporates these changes. It can be used within the IEEE 802.1x standard. During every authentication exchange, the Home Authentication Server (HAS) randomly generates a new temporary client identifier, so that new and old identifier are unlinkable for eavesdroppers. Apart from the client identifier, EAP-TLS also transmits a client certificate, which could provide a means for tracking because it remains constant during every authentication. The modified protocol encrypts the certificate with the authentication server's public key and a randomly chosen salt, so that the exposed bitstring changes for every transaction.

Moreover, the implementation must ensure that the protocol does not advertise a unique machine configuration. For example, the TLS client_hello message contains the system time and a list of supported encryption standards. The implementation could add a random drift to the TLS clock and report only the most common encryption standard configuration.

To further illustrate the application of these privacy techniques, Table 2 describes a CHAP-derived protocol, which does not require a public key infrastructure. It assumes that the client and Home Authentication Server share a symmetric secret key k, which is exchanged at the time of subscription. The protocol comprises a two-way challenge-response handshake between the client and the HAS.

To protect against active attacks from the Foreign Authentication Server, the client should only accept messages that include the correctly encrypted random challenge. The Home Authentication Server should store the current and previous temporary identifiers for each client, in case the message containing the new identifier is dropped (maliciously or accidentally).

3.5. Implementation issues

A new address is selected before the client associates with the access point; for example, after booting or waking up from standby. At this time no disruptions occur since no network connections can exist. Furthermore, signal strength parameters have likely changed; this prevents correlation of new and old address through similar signal strength measurements.

We prototyped the MAC address switch through the following `ifconfig` call.

```
ifconfig eth0 hw ether 01:02:03:04:05:06
```

This procedure was successful on a Linux 2.4 kernel using either a Dell 4800 LT card with the aironet driver or a Linksys WPC11 v2.5 card with the hostAP driver.

[4] This assumes that every network has a sufficiently large number of clients, otherwise the network identifiers could become pseudo-identifiers for individual clients.

Table 1
Privacy-preserving EAP-TLS-derived authentication protocol between Supplicant (S) and an Authenticator (A).

1.	A requests identity from S	EAP-Request/Identity
2.	S answers with network and temporary client ID	EAP-Response/Identity {HNID, TCID}
3.	A forwards {HNID, TCID} to FAS	
4.	FAS forwards the request to the HAS specified by the HNID	
5.	HAS verifies that TCID is valid and starts TLS authentication	TLS Start
6.	S sends hello to HAS	TLS client_hello
7.	HAS supplies its own certificate and requests S's certificate	TLS server_hello, certificate, server_key_exchange, certificate_request, server_hello_done
8.	S sends a public key encrypted certificate	E_{pks}{salt, client_certificate}, client_key_exchange, certificate_verify, change_cipher_spec, finished
9.	HAS sends new client identifier	$E_{\text{pkc}}\{TCID_{\text{new}}\}$,TLS change_cipher_spec, finished
10.	S acknowledges	EAP-Response
11.	HAS reports the result of certificate verification and instructs the AP to block or permit access	EAP-success/EAP-Failure

The protocol uses the roaming network's Foreign Authentication Server (FAS) and the subscriber's Home Authentication Server (HAS). The second column shows the messages received/sent by the supplicant. The home network identifier (HNID) refers to the client's home service provider. The temporary client identifier (TCID) is exchanged between S and HAS, and E_{pks}/E_{pkc} means encrypt with the public key from server/client.

Table 2
Privacy-preserving authentication between Client (C) and an Access Point (AP) using the roaming network's Foreign Authentication Server (FAS) and the subscriber's Home Authentication Server (HAS).

1.	C sends association request to AP {HNID, TCID, R_c}
2.	AP forwards the request to FAS
3.	FAS forwards the request to the HAS specified by the HNID
4.	HAS uses TCID to look up the subscriber's key and responds with $\{E_k(R_c, TCID_{\text{new}}, R_a), R_a\}$
5.	The FAS forwards $\{E_k(R_c, TCID_{\text{new}}, R_a)\}$ and stores the remainder
6.	C checks the response to his challenge, stores the new TCID, and responds to HAS challenge with R_a
7.	FAS checks the client's response and instructs the AP to block or permit access.

The home network identifier (HNID) refers to the client's home service provider. The temporary client identifier (TCID) is exchanged between C and HAS, R stands for random challenges, and E_k means encrypt with the shared key k.

Unfortunately, the 802.11 link layer protocol further complicates the implementation of unlinkable identifiers. It includes sequence numbers in the frame header. If these sequence numbers are maintained after an address switch, the adversary could exploit these numbers to identify the previous address of a user. Therefore, it is necessary to reset or unpredictably change the sequence number. Power cycling a PCMCIA card achieves this goal; however, it adds a significant delay. We are investigating alternative techniques for changing the sequence numbers.

4. Privacy evaluation

WLAN installations differ in their configurations and in the environments, in which they are deployed. We identify several factors that affect the degree of location privacy enjoyed by the networks users.

*Open vs. closed environment***.** Open environments are characterized by a high fluctuation among clients; that is, the adversary does not know possible client interface identifiers in advance. An example of a closed environment is a company network, where all authorized client's interface identifiers are registered. Thus, if an address appears that is not registered, the adversary can in infer that the address has been spoofed.

*Frequency of address switches***.** In a network where only a single user installed the address switching system, the adversary can more easily track this user. The adversary only needs the ability to distinguish an address switch from a client that newly associated with the network. In a network with multiple address-switching clients, such an adversary would be unable to distinguish between them.

*Location resolution***.** Wireless LAN installations typically contain multiple access points that together serve a larger area. Location resolution describes how accurately an adversary can locate a client within this service area. Assuming that clients are not continuously moving, more accurate location information would help the adversary in linking the addresses before and after a switch.

*Prior knowledge***.** Knowledge about the environment such as building layout or office assignments contributes to user identification. For example, a signal originating from a particular office would identify the owner of this office as the user of the received MAC address. If no address switches occur, this would enable the adversary to link the past and future movements of this MAC address to the user.

4.1. Quantitative analysis

The system is designed to reduce the time during which a user can be tracked. Even if an adversary can identify the user of a certain interface identifier, the adversary probably loses track of this user when the next identifier switch occurs. For example, the adversary could recognize that the wireless signal originates from a certain office. This would likely reveal the identity of the user. However, when the user leaves his office and the MAC address switches, the adversary would have difficulties to determine whether the new MAC address belongs to the same or another user.

An identifier switch does not guarantee that the adversary will be unable to track the user. The adversary potentially analyses other clues to find a correlation between different addresses. Assuming a static setting, where clients move infrequently, the adversary can link new and old addresses using signal characteristics. Through an address switch, the signal strength and signal-to-noise ratio properties of a client remain unchanged, unless the client moves. Although these properties might fluctuate, the adversary could filter out noise and infer which address belong to the same client.

Overall, the effectiveness of address switching, with respect to protecting location privacy and minimizing network disruptions, depends on the usage characteristics of the WLAN. For instance, a high rate of associations and disassociations in the network will better disguise the address switching of a single client.

Therefore, the quantitative analysis is designed to answer these key questions:

– How long is the expected tracking time when switching addresses on association with an access point?
– Is there a sufficient user population in public WLANs?
– How vulnerable is address switching to more sophisticated analysis attacks?

The last question is most difficult to answer since an adversary could – assuming a more substantial effort – exploit a vast range of information from physical layer signal characteristics, over packet interarrival times, to cookies. We limit the scope of the last question to signal-to-noise (SNR) information since it is readily available (to an adversary) from WLAN access points and it is specific to the wireless medium.

The analysis is based on usage traces from a public WLAN. Specifically, we chose a trace from a IEEE 802.11b installation covering the auditorium and lobby area during the 2001 IEEE SIGCOMM conference in San Diego [4]. The trace was collected through the Simple Network Management Protocol (SNMP) interface of 4 Orinoco AP-100 access points at approximately one-minute intervals over 52 hours during the conference. Among others, the data contain at each access point the MAC addresses, the signal-to-noise ratio (SNR), and packet counts for associated clients. We chose the conference setting over more extensive traces taken on a university campus [22] and in a corporate environment [5], because user behavior at a conference better reflects the characteristics of a public hotspot.

From the MAC address lists, we compute the association time for each session. The signal-to-noise ratio data in the trace enable the approximate simulation of a more sophisticated adversary that exploits signal characteristics. This algorithm keeps track of the addresses and their last mean SNR from clients that have disassociated at each access points. When a new client associates, the algorithm looks for a SNR match in its table. If an entry matches, the adversary assumes that the new address belongs to the same client that has been observed previously; otherwise, the client remains unidentified.

4.2. Results

Figure 3 shows a histogram of the association times (per session) measured in the conference setting. Notice that the time axis is plotted on a logarithmic scale. Most sessions are relatively short – below 10 minutes. This probably indicates users who briefly check their email. However, the variance in association times is very high. A fair number of sessions lasts between 30 minutes and 3 hours and a few clients remain connected for virtually a full day. Therefore, our system offers good protection for the majority of clients. A few, however, would benefit from mechanisms that could switch addresses more frequently.

The distribution of clients across the 4 access points is illustrated in figure 4. The graph shows the Tukey plot (min, max, median and inter-quartile range) of the number of simultaneously associated clients at each access point. Apart from slight variations, the clients are virtually uniformly distributed across the four access points. The median and minimum lie at approximately 5 clients. Again, this shows that a small number of clients remain associated throughout the night. The third quartile varies between 14 and 17 clients across the access points. The maximum is reached at 30 to 35 clients per access

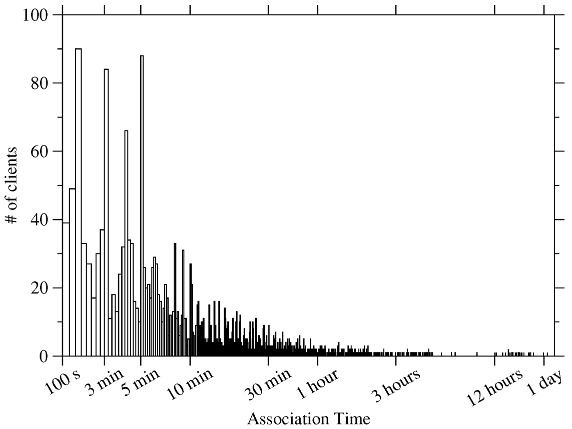

Figure 3. Distribution of client association time. The majority of clients stays associated for less than 10 minutes, while a few clients use the same address for over half a day. The time axis is plotted on a logarithmic scale.

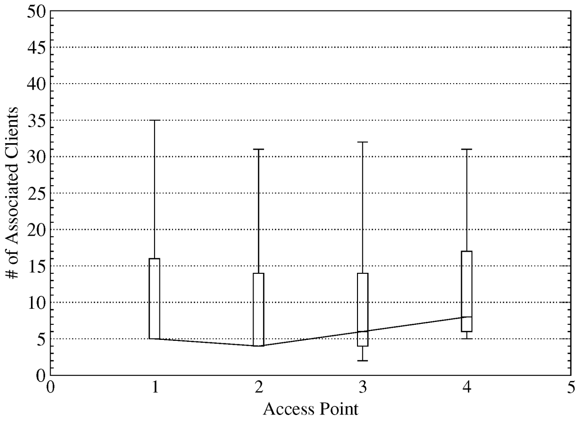

Figure 4. Tukey plot (min, max, median and inter-quartile range) of the number of simultaneously associated clients at four different access points.

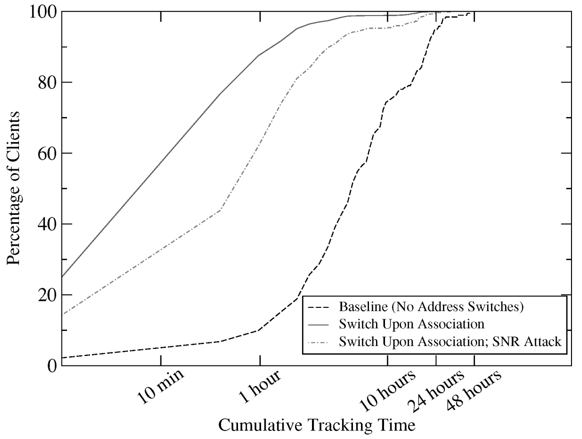

Figure 5. Normalized cumulative distribution functions (CDF) showing the tracking time for each identifier. Addresses switches on association significantly reduce the time during which an identifier can be tracked. The time axis is plotted on a logarithmic scale.

point. A sufficient number of clients are present during the day so that associations occur frequently and identities are not obvious.

Figure 5 shows how long an adversary can track identifiers. Tracking time is plotted as a normalized cumulative distribution function with a logarithmic x-axis; that curve indicates the fraction of the sample population (y-axis) that can be observed for no more than a specified duration (x-axis). The "baseline" curve describes an unprotected system, where no address changes occur. Thus, the tracking time shows the total time over multiple sessions during which a client is associated. In this case, the adversary can track about 90% of clients for more than 1 hour (total time over multiple sessions) and 25% of clients for more than 10 hours out of the 52 hours of the trace. Switching the interface identifier upon association with an access point significantly reduces the tracking time; less than half the pseudonyms are tracked for more than 10 minutes and only 10% are observed for more than an hour. However, the adversary can improve the tracking performance through correlating pseudonyms with similar SNR characteristics. This technique yields 40% of pseudonyms tracked for more than an hour.

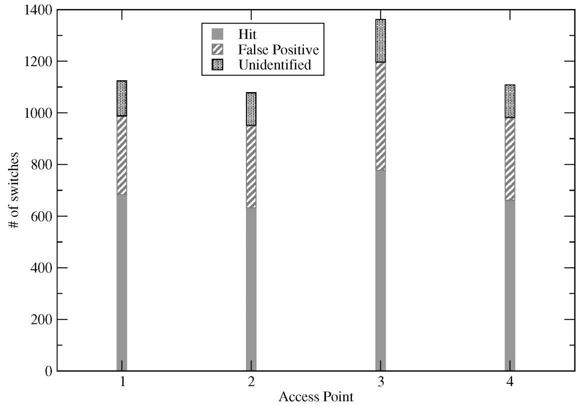

Figure 6. Effectiveness of signal-to-noise ratio correlation attack. Linking identifiers succeeds in about half the attempts but also suffers from a significant false positive rate.

The effectiveness of a more capable adversary using the SNR technique to link addresses is further illustrated in figure 6. For each of the 4 access points, the bar describes the number of correctly linked switches (hit), the number of incorrectly linked switches (false positive), and the number of switches that remained unlinked. The hit rate lies consistently between 50% and 60% but the high rate of false positives reveals that such tracking results are unreliable. This initial result looks promising; however, we note that more sophisticated techniques such as triangulation from several access points, or correlation using multiple factors (e.g., HNID field, SNR, network traffic patterns) should be analyzed in future work.

5. Discussion

While the initial results look promising, we discuss further challenges in increasing privacy for clients with long associations, and integrating with access control mechanism and encryption.

5.1. Network disruptions

When switching addresses only on association is insufficient, network disruption problems arise. An address switch while associated would abort all network connections (e.g., TCP connections), because the client also needs to obtain a new IP address. Thus, switching requires proper timing, not only to preserve privacy, but also to minimize disruptions.

The appropriate time could be determined through several mechanisms. After a random time interval has passed the system could detect periods without open network connections and use these to reassociate the client with a different address. Some users may even allow the system to drop connections for a higher degree of privacy. This should be configurable through user preferences. In addition, the address switch could be triggered by significant variations in signal strength. At this time, the access points will probably notice similar fluctuations in signal strength and find it more difficult to track the client.

Another interesting approach is spoofing multiple clients. It is possible to use a wireless network card to send different beacon frames which advertise an access point [8]. This method essentially spoofs a large number of access points.

It would be useful if such approaches could be extended so that one wireless card can maintain an association to an access point and communicate on *multiple MAC addresses simultaneously*. This would enable a client to spoof other clients in the vicinity and thus increase privacy. It would also smooth the transition between network addresses, because a client could open connections on a new address while maintaining other connections on the old address. However, we fear that it will be difficult to control time-sensitive operations such as transmitting link layer acknowledgements outside the firmware.

5.2. Access control mechanisms

The described solution conflicts with MAC-based security measures that aim to protect a private wireless network from unauthorized intruders. Some access point operators implement access control mechanisms that restrict network service to a list of authorized MAC addresses. Obviously, an authorized client would be denied access if the client presents a randomly selected MAC address, which is not on the list. At least two possible approaches address this conflict.

First, for networks that employ MAC address access control, it may be possible to scan the network with a client adapter to build a list of valid MAC addresses. When the detected cards become inactive, these addresses could be used to achieve at least a small degree of privacy. Further study is needed to determine the feasibility of this approach.

Second, other access control mechanisms do not necessarily conflict with interface identifier disposal. For example, authorization based on a shared WEP key preserves a high degree of anonymity, because every authorized party uses the same key, and allows access from any MAC address. Some private networks currently do not use MAC address access control, because other mechanisms such as WEP are more secure. In other networks, a cooperating authority[5] could explicitly allow the usage of privacy-enhancing technologies, such as the authentication protocols described in Section 3.4.

5.3. Encryption

Integration of identifier disposal with encryption would be mutually beneficial. Identifier disposal can enhance content privacy (provided that the messages contain no identifying information), since encryption alone does not provide complete protection. Link layer encryption does not protect against untrustworthy access points and transport layer encryption does not hide the network addresses (which identify the communication partners). While several specialized mechanisms for anonymous network connections have been devised (e.g., [16,31]), we believe that a major advantage of interface identifier disposal is that it provides increased anonymity and thus content and communication privacy in situations where the specialized techniques are not available

Encryption also strengthens identifier disposal since it hides message content from potential adversaries. That is, this information is not available for traffic analysis attacks.

5.4. Billing

The authentication scheme for roaming scenarios described here may need to provide additional security features. These must also be implemented in a privacy-preserving manner. For example, commercial service providers may also need to meter service usage for billing purposes. In roaming scenarios, foreign access point operators may try to deprive a paying client of bandwidth or exaggerate the duration of service usage; thus, network operators and users have an interest in an integrity-preserving billing system.

How can the home authentication server measure the service time used by the client, when neither the client nor the access point are fully trustworthy? An intuitive approach is to reauthenticate the client after a time interval t has elapsed, say one minute, and bill the client for each started interval. To reduce the communication load on the home authentication server, the foreign authentication server could require the supplicant to sign a lease for each interval of service used. These can be processed offline by the home authentication server.

While these mechanism improve billing integrity, they do not completely address billing fraud by foreign authentication servers. Even if the FAS denies service for the agreed upon time interval, the FAS can still bill the client. If the time interval is sufficiently short, this may be a small risk for an individual client; however, it still adds up to a significant value when the FAS repeats this scheme with a large number of clients. The complete solution remains an open problem.

6. Related work

Location privacy risks have been discussed in a variety of technologies. For IEEE 802.11b WLAN systems, Smailagic and Kogan [33] describe these risks in the context of an WLAN-based position sensing system and we have assessed the location privacy problems in public WLANs [18]. Smailagic and Kogan addresses privacy concerns through a client-centric location sensing mechanisms, so that no potentially untrustworthy servers need to be involved. However, this does not prevent a malicious access point operator to obtain location information from the access points directly (unless the client

[5] Notice that this authority does not have to be the only one with control over the access points. For example, in a corporate environment higher-level management could support the deployment of privacy-enhancing technology to protect their employees from malicious technicians and insiders. Thus, the assumption of a cooperative authority with policy control over the access points does not imply that access points are trustworthy.

never transmits a signal; that is, it is not used as a network interface).

The Mist routing project for mobile users [2] combines location privacy with communication aspects. It focuses on the problem of routing messages to a subject's location while keeping the location private from the routers and the sender. To this end, the system is comprised of a set of mist routers organized in a hierarchical structure. The leaf nodes have knowledge of the location of users but not their identity. They refer to them through handles (or pseudonyms). Each user selects a higher-level node in the tree, which acts as a semi-trusted proxy. It knows the identity of the user but not his exact location. The paper then presents a cryptographic protocol to establish connections between users and their semi-trusted proxies and mechanisms to connect to communication partners through their proxies. The paper does not address the location privacy problems caused by unique addresses at the IEEE 802.11b link layer.

For Mobile IP, Fasbender et al. [15] propose the Non-Disclosure-Method (NDM), which reroutes traffic to hide a mobile user's location. Mobile IP enables hosts to transparently migrate between different networks. When the host moves away from his home network, a home agent receives the traffic and tunnels it to a foreign agent in the host's new network. This requires that the host registers the care-of address, his current location, with the home network. Thus, an adversary can track the location of a host by observing the registration messages and the payload messages through the tunnel. The NDM method places several security agents between home and foreign agent. The security agents forward messages in encrypted form; therefore, it is hard to trace the path of a message if the security agents are spread over several administrative domains. This method hides the network layer care-of network address of a host from intermediary routers, but it does not protect the interface identifier from adversaries on the local area network.

Narten and Draves propose privacy extensions for stateless address autoconfiguration in IPv6 [27]. Server-less address autocofiguration in IPv6 can use the MAC address of network interfaces as part of the network layer IP-address. Thus the MAC address becomes visible to servers and intermediaries outside the local area network. This enables such outside entities to track movements of mobile nodes between different networks. The proposed solution switches the network address periodically. New addresses are generated through iterative applications of the MD5 message digest algorithm on the previous network address and the actual MAC address. Again, this approach assumes that the local area network is trustworthy; thus, it concentrates on protecting the address from parties outside the LAN.

Location privacy enhancements for GSM mobile phone networks seek to prevent foreign network operators from obtaining the subscriber identity [21,24]. This work assumes a trustworthy home network operator, because the subscriber enters into a contractual agreement with this party. However, the identity information should be protected from other network operators that could learn the subscriber identity and location in roaming scenarios. To this end, temporary pseudonyms are generated, that only the home network operator can map to the real identity. This research is similar in goal, it protects user identity from unknown network operators. Our approach differs in its design for IEEE 802.11b networks and addresses the resulting problems, such as address selection without a trusted agent. Furthermore, we quantitatively analyze the privacy protection offered by this system.

Several researchers designed privacy-enhancing location servers [17,25,34] for situations where location data needs to be revealed to external services. Specifically, the IETF *Geopriv* working group is addressing privacy and security issues regarding the transfer of high resolution location information to external services and the storage at location servers. It focuses on the design of protocols and APIs that enable devices to communicate their location in a confidential and integrity preserving manner to a location server. The location server can reduce the data's resolution or transform it to different data formats, which can be accessed by external services if the data subject's privacy policy permits. The working group is also interested in enabling unidentified or pseudonymous transfer of location information to the server and access from the server. Such mechanisms targeted at the application layer provide no protection at the link layer.

Address conflicts have been investigated both at the network and link layer. For example *gratuitous ARP* [35, p. 62] detects duplicate IP addresses on the network. Conflicts are usually manually resolved. Similarly, MAC address conflicts can occur through manufacturing errors or locally configured MAC addresses. Symptoms and procedures for manually correcting such problems are described in the Microsoft Knowledge Base [6]. More recently, the challenge of assigning network and hardware addresses has arisen in ad-hoc networks. Addresses must be configured in environments where dedicated infrastructure, such as a DHCP server, is not present. In addition, frequent network partitions complicate duplicate address detection. Vaidya [37] proposes a modified routing and duplicate address detection scheme, that can tolerate network address conflicts, as long as packets are always delivered to the node intended by the sender. Schurgers and colleagues [32] and Bharghavan [7] reduce the overhead from large static hardware addresses through shorter dynamically assigned addresses in wireless sensor networks and cellular WLAN, respectively. However, the assignment algorithms rely on all nodes in the network implementing the distributed assignment algorithms. Furthermore, they do not indicate how addresses can be assigned in a privacy-preserving manner (so that new and old address cannot be linked).

7. Conclusions and future work

In this paper, we have considered the threat model and solutions for location privacy in wireless 802.11 networks. We have proposed MAC rekeying as a solution and analyzed the situations in which it may apply. We have shown that in large networks, address collisions can become a problem, which means that simply selecting a random MAC address is insufficient and duplicate address detection mechanisms need to be redesigned from a privacy perspective. The rekeying mechanism can be integrated with authentication protocols and, for most hosts, substantially reduced the tracking time – as observed in the wireless network of a recent conference – without network disruptions. Furthermore, in crowded public networks, identification attacks based on signal characteristics first need to overcome the high error rate that we have seen based on analysis of a large public wireless network.

We believe that as wireless networks become more ubiquitous, concerns about location privacy will heighten. The techniques analyzed in this paper are in tension with current widely-deployed security and access control mechanisms which rely on a unique MAC address for authentication. Emerging standards should be designed with location privacy concerns in mind so that they do not require a constant unique identifier that is exposed on the wireless medium.

In future work, we plan to analyze the system's protection against more sophisticated adversaries. Using several access points, an adversary could locate a client more accurately and use this position information instead of single SNR measurements to link addresses to the same client. In addition, traffic analysis techniques should be taken into account. We also plan to investigate mechanisms that can unobtrusively switch addresses while a client is associated. This would offer additional protection during longer sessions. Furthermore, usage characteristics and their privacy implications should be investigated in environments other than a conference setting.

Acknowledgments

John Black provided valuable feedback and helped with the birthday paradox. This work was supported by research contracts from the National Science Foundation.

References

[1] P.E. Agre, RRE notes and recommendations (1999) http://commons.somewhere.com/rre/1999/RRE.notes.and.recommenda14.html.

[2] J. Al-Muhtadi, R. Campbell, A. Kapadia, M.D. Mickunas and S. Yi, Routing through the mist: Privacy preserving communication in ubiquitous computing environments, in *International Conference of Distributed Computing Systems* (2002).

[3] P. Bahl and V.N. Padmanabhan, RADAR: An in-building RF-based user location and tracking system, in: IEEE INFOCOM (2000) pp. 775–784.

[4] A. Balachandran, G. Voelker, P. Bahl and P. Rangan, Characterizing user behavior and network performance in a public wireless LAN, in: *Proceedings of ACM SIGMETRICS* (2002).

[5] M. Balazinska and P. Castro, Characterizing mobility and network usage in a corporate wireless local-area network, in: *The First International Conference on Mobile Systems, Applications, and Services (MobiSys)* (2003).

[6] M.K. Base, Article 164903 – How to troubleshoot duplicate media access control address conflicts, (2002) http://support.microsoft.com/default.aspx?scid = KB;en-us;q164903.

[7] V. Bharghavan, A dynamic addressing scheme for wireless media access, in: *International Conference on Communications* (1995).

[8] Black Alchemy, FakeAP (2003) http://www.blackalchemy.to/project/fakeap/.

[9] N. Borisov, I. Goldberg and D. Wagner, Intercepting mobile communications: the insecurity of 802.11, in: *Proceedings of the Seventh Annual International Conference on Mobile Computing and Networking* (2001) pp. 180–189.

[10] P. Castro, P. Chiu, T. Kremenek and R. Muntz, A probabilistic room location service for wireless networked environments, in: *Ubicomp* (2001).

[11] CNN, Police: GPS device used to stalk woman, (2002) http://www.cnn.com/2002/TECH/ptech/12/31/gps.stalk.ap/index.html.

[12] EAP TLS, PPP EAP TLS authentication protocol Requests for Comments 2716 (1999).

[13] D. Eastlake, S. Crocker and J. Schiller, RFC 1750: Randomness recommendations for security, (1994) http://www.ietf.org/rfc/rfc1750.txt.

[14] A. Ebner and H. Rohling, A self-organized radio network for automotive applications, in: *Proceedings of the 8th World Congress on Intelligent Transportation Systems* (2001).

[15] A. Fasbender, D. Kesdogan and O. Kubitz, Analysis of security and privacy in mobile ip, in: *4 th International Conference on Telecommunication Systems Modeling and Analysis* (1996).

[16] D. Goldschlag, M. Reed and P. Syverson, Onion routing for anonymous and private Internet connections, Communications of the ACM (USA) 42(2) (1999) 39–41.

[17] M. Gruteser and D. Grunwald, Anonymous usage of location-based services through spatial and temporal cloaking, in: *Proceedings of the First International Conference on Mobile Systems, Applications, and Services* (2003a).

[18] M. Gruteser and D. Grunwald, A methodological assessment of location privacy risks in wireless hotspot networks, in: *Proceedings of the First International Conference on Security in Pervasive Computing* (2003b) (to appear).

[19] IEEE, 1999, IEEE Standard 802.11b – Wireless LAN Medium Access Control (MAC) and Physical layer (PHY) specications: High Speed Physical Layer (PHY) in the 2.4 GHZz Band (1999).

[20] IEEE, OUI assignments (2003) http://standards.ieee.org/regauth/oui/index.shtml.

[21] D. Kesdogan, H. Federrath, A. Jerichow and A. Pfitzmann, Location management strategies increasing privacy in mobile communication, in: *12th International Information Security Conference*. Samos, Greece (1996) pp. 39–48.

[22] D. Kotz and K. Essien, Analysis of a campus-wide wireless network, in: *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking* (2002) pp. 107–118.

[23] A.M. Ladd, K.E. Bekris, A. Rudys, L.E. Kavraki, D.S. Wallach and G. Marceau, Robotics-based location sensing using wireless ethernet, in: *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking* (2002) pp. 227–238.

[24] C.-H. Lee, M.-S. Hwang and W.-P. Yang, Enhanced privacy and authentication for the global system for mobile communications, Wireless Networks 5(4) (1999) 231–243.

[25] U. Leonhardt and J. Magee, Security considerations for a distributed location service, Journal of Network and System Management 6 (1998) 51–70.

[26] Location Privacy Protection Act, Location privacy protection act, (2001) http://www.techlawjournal.com/cong107/privacy/ location/s1164is.asp.

[27] T. Narten and R. Draves, RFC3041 – Privacy Extensions for Stateless Address Autoconfiguration in IPv6, http://www.faqs.org/ftp/rfc/rfc3041.txt.

[28] N. Negroponte, Being wireless, Wired 10(10) (2003).

[29] M. Piszczalski, The next big thing: Wi-Fi, Automotive Design and Production (2002).

[30] J. Reed, K. Krizman, B. Woerner and T. Rappaport, An overview of the challenges and progress in meeting the e-911 requirement for location service, IEEE Personal Communications Magazine 5(3) (1998) 30–37.

[31] M.K. Reiter and A.D. Rubin, Crowds: Anonymity for Web transactions, ACM Transactions on Information and System Security 1(1) (1998) 66–92.

[32] C. Schurgers, G. Kulkarni and M.B. Srivastava, Distributed assignment of encoded MAC addresses in sensor networks, in: *Proceedings of the 2001 ACM International Symposium on Mobile Ad Hoc Networking and Computing* (2001) pp. 295–298.

[33] A. Smailagic and D. Kogan, Location sensing and privacy in a context-aware computing environment, IEEE Wireless Communications 9 (2002) 10–17.

[34] M. Spreitzer and M. Theimer, Providing location information in a ubiquitous computing environment, in: *Proceedings of the Fourteenth ACM Symposium on Operating System Principles* (1993) pp. 270–283.

[35] W.R. Stevens, *TCP/IP Illustrated*, (Addison-Wesley, 1994), Vol. 1.

[36] A. Stubblefield, J. Ioannidis and A. Rubin, Using the Fluhrer, Mantin, and Shamir attack to break WEP, Technical Report TD4ZCPZZ, ATT Labs (2001).

[37] N.H. Vaidya, Weak duplicate address detection in mobile ad hoc networks, in: *Proceedings of the Third ACM International Symposium on Mobile Ad Hoc Networking and Computing* (2002) pp. 206–216.

[38] Wireless Geographic Logging Engine, Wireless geographic logging engine (2002) http://wigle.net/gpsopen/gps/GPSDB/.

Marco Gruteser is a Ph.D. candidate in computer science, advised by Prof. Dirk Grunwald at the University of Colorado at Boulder. His research interests include location privacy, context-aware applications, and wireless networks. He received his MS in computer science from the University of Colorado at Boulder and completed a Vordiplom at the Technical University Darmstadt, Germany. During a one-year leave at the IBM T.J. Watson Research Center, he developed software infrastructure that integrates sensors to support context-aware applications in the BlueSpace smart office project. He is a student member of the ACM. Contact him at Campus Box 430, Boulder, CO 80309-0430;
E-mail: gruteser@colorado.edu; http://www.cs.colorado.edu/ gruteser.

Dirk Grunwald received his Ph.D. from the University of Illinois in 1989 and joined the University of Colorado the same year. His work addresses research and teaching in the broad area of “computer systems”, which includes computer architecture, operating systems, networks, and storage systems. His interests also include issues in pervasive computing, novel computing models, and enjoying the mountains. He is currently an Associate Professor in the Department of Computer Science and in Electrical and Computer Engineering and is also the Director of the Colorado Center for Information Storage.