Privacy and Security in the Location-enhanced World Wide Web

Jason I. Hong[1], Gaetano Boriello[2,3], James A. Landay[2,3], David W. McDonald[4], Bill N. Schilit[2], J. D. Tygar[1]

[1] Computer Science Division, University of California, Berkeley
[2]Intel Research Seattle, 1100 NE 45th Street, Seattle, WA 98105
[3]Computer Science & Engineering, University of Washington
[4]Information School, University of Washington

ABSTRACT

Privacy concerns remain a major barrier to adoption of location-based services. Users demand significant, concrete benefits before they are willing to allow an outside party to track their movements. We propose **Place Lab,** a trustworthy, secure location infrastructure that gives users control over the degree of personal information they release. Place Lab allows service providers to design services that provide basic benefits to all users, and expanded benefits as users release personal location information. This position paper discusses privacy and security issues arising in Place Lab, a multi-organization initiative to bootstrap the location-enhanced web.

Keywords

Location-aware, context-aware, WiFi hotspot, World Wide Web, privacy, security.

INTRODUCTION

Location-aware computing has been a major focus of ubiquitous computing research since its inception. Location helps organize information and services as well as contextualizes people's activities. However, with the possible exception of GPS-based navigation, location-aware applications are rarely used by the general population, despite advances by the research community. We believe there are at least three challenges that must be met before the widespread adoption of location-aware computing, specifically:

- low-cost, convenient location finding technologies;
- useful, usable location-based services; and,
- techniques to address end-user concerns about location privacy (the topic of this paper).

The goal of Place Lab is to provide an open software base and a community building activity for overcoming these barriers and catalyzing adoption of location-aware computing by end-users and service providers. Place Lab uses the wide deployment of 802.11b WiFi access points (APs) for determining one's location. A key observation here is that many developed areas have wireless hotspot coverage so dense that cells overlap. By consulting the Place Lab directories, which will map wireless hotspot MAC addresses to physical locations, mobile computers and PDAs equipped with WiFi can determine their location to within a city block[1]. Since WiFi is widely used across a broad variety of mobile devices, this positioning technology is extremely low-cost. Furthermore, by keeping cached copies of these directories, computers can calculate their location locally *without transmitting information to any other computer*. At this basic level, if one's personal device does not use the same wireless hotspot for communication, then this technique yields a totally private positioning technology. In these cases, one can still interact with cached offline content in an occasionally connected computing (OCC) model. For example, if the Zagat restaurant guide[2] was an OCC location-enhanced site, an end-user could get information about nearby restaurants without revealing any location information to Zagat's web servers.

However, we expect that there will be many useful services that cannot be made available in an OCC model, such as real-time information describing current weather and traffic conditions, mobile commerce techniques that allow interaction between potential merchants and consumers, and other multi-user interactive applications. In these cases, we expect there would be some user interface features that make it easy for end-users to share their location information at a level they are comfortable with.

Place Lab is currently under development. In a previous paper we discussed the overall vision of Place Lab and provide more details about its design [1]. This position paper analyzes privacy issues in Place Lab. Specifically, we identify various stakeholders, evaluate the potential threats to these stakeholders, and examine how they are managed by Place Lab. We hope that this will serve as a starting point for discussions about privacy in Place Lab.

[1] MAC addresses are globally unique IDs for both WiFi cards and WiFi access points. MAC addresses for APs are also known as BSSIDs (Basic Service Set Identification), and are broadcast by access points to all nearby receivers — there is no interactive exchange necessary. Access points also broadcast an SSID, which is the name of the wireless network. It should be noted that some checking is necessary — MAC addresses can be forged, and access points can be moved — but the vast majority of AP MAC addresses are unique and statically located.

[2] http://www.zagat.com/

LOCATION PRIVACY ISSUES

Privacy means different things to different people in different situations. For example, a clear glass window is as good for looking into spaces as it is for looking out. Staring into a local department store window is acceptable behavior, but staring into my neighbors windows is socially unacceptable and often illegal [2]. We need to represent privacy concerns in a concrete way to address them.

Concepts of privacy for location-aware technologies co-evolve with the technologies. Location-aware technologies are locked into this socio-technical space because many of the social concerns are socially and contextually constructed. Successful location-aware technologies will not simply recognize this fact, but will facilitate and enable the co-evolution of both the application as well as the social attitudes which develop around the system.

To illustrate these points, it is useful to examine specific privacy issues faced by previous ubiquitous computing systems. Below, we discuss three privacy issues encountered by an early ubiquitous computing system, the PARCTab system [3].

First, PARCTab used a centralized server to hold location data. While this architecture made it easier to create certain kinds of applications, it meant that sensitive data was stored on a computer that end-users had little practical control over. Even though a visible effort was made to create written privacy policies, users still had the perception that if the research team managing the system changed their policies, or if upper-level managers wanted to examine the data, there was little they could do about it. In addition, centralized servers are attractive targets for computer security attacks.

Second, there was no control over the level of location information disclosed. By design, PARCTab base stations continuously forwarded location information to higher level processes. Even without running applications, the device's location was known because it beaconed a data packet for this purpose. The system was "all or nothing": users did not have any granular control over the degree of information sent (it specified location by room) or whether that information was shared with others. There were no provisions for ambiguity or for tailoring the level of disclosure to suit individual preferences.

Third, there was no disclosure over what information was revealed to third parties. A stranger could monitor a user's location by making repeated queries about the user's location without that user knowing.

To summarize, the PARCTab system exposed three significant privacy issues:

- centralized architectures require users to trust the operators of the service, both to properly use location data and to sufficiently protect it;
- end-user control over location data should provide more granularity than a binary on- or off-switch, and should allow control over when, to whom, and how much location information is shared;
- users want to know when and to whom systems share their user location information.

MANAGING PRIVACY IN PLACE LAB

Over the past few years, the research community has been moving from centralized architectures for maintaining location data to decentralized ones. In centralized architectures such as Active Badge [4], Active Bats [5], and PARCTab [3], the infrastructure consists of receivers deployed in places of interest, with end-users beaconing out data stating that "I am here." One's location data is initially determined on computers outside of one's personal control. In contrast, in decentralized architectures such as Cricket [6] and RADAR [7], the infrastructure consists of beacons deployed in places of interest, signaling to end-users that "You are here." In decentralized architectures, one's location is initially determined on a personal device, giving end-users greater choice over whether to disclose their location to others and what information is disclosed.

Place Lab uses the decentralized approach, relying on WiFi hotspots as beacons. In this architecture, mobile notebook and PDA computers detect access points and then look up the access point's MAC address in a local directory of hotspots. In cases where users can detect multiple hotspots, the intersection of their coverage can be used to calculate more precise location. Users might take this location data and use it locally with applications such as MapPoint[3], or they might connect to the Internet and send their location to web services. The important consequence is that users can trade privacy for utility on a case-by-case basis, much as they decide on a case-by-case basis whether to enter a credit card or phone number when asked by a web site.

Client-computed location is the foundation for the most flexible privacy mechanisms and policies. In this model the stakeholders groups include: (1) end-users; (2) access point owners; (3) network service providers; (4) web service providers. In the following section we examine some of the issues that relate to each of these entities.

End-User Privacy

Privacy for end-users is the most complex part of the model, as it involves interactions with all of the different stakeholders. It is useful here to separate end-user interaction into disconnected and connected cases:

The *disconnected* case is simple, since end-users use access points only to calculate their position. This can be done passively, without revealing any information to access points[4]. No information is transmitted to others.

In the *connected* case, location can still be calculated locally, but now information is transmitted through an

[3] http://mappoint.msn.com/

[4] For example, Airsnort (http://airsnort.shmoo.com) does this to crack WEP keys.

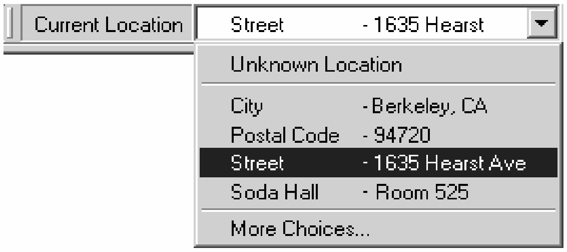

Figure 1: When location is computed locally, how might users manage what is revealed to external hosts? The *Place Bar* browser component (mockup) lets users select the level of location information disclosed to web sites, potentially on a page by page basis, as part of their browsing activity.

access point, across a network service provider, and to a web service provider.

We envision that privacy can be managed by end-users at the user interface level through the *Place Bar*; a component integrated into web browsers that makes it easy for end-users to selectively reveal their location at various granularities to location-enhanced web sites (see Figure 1). The Place Bar provides control over and feedback about one's location in a single, simple interface. For example, an end-user could choose to reveal their location at the city level to retrieve current events going on in that city, or at the street level to locate nearby drugstores. End-users can selectively trade the privacy of their location for increased utility of online content.

This does not mean, however, that all location information flows must be manually configured. Permissions could also be managed on automated criteria. For example, when you are in the office, perhaps you want your location to be generally disclosed. Another example: you may have certain location-enhanced web services you trust. The Place Bar provides a simple and convenient mechanism for managing the release of one's current location information to location-enhanced web services.

When users transmit information to an open access point, they do reveal their MAC addresses. MAC addresses can be difficult to map back to specific users, but it may be possible to infer the user's identity if access points share information or if a user repeatedly transmits in the vicinity of an access point. To address this concern, it may be appropriate to choose rapidly changing MAC addresses to prevent correlations that yield user identities.

Note that anyone can scan the content being transmitted over open access points, so this analysis only holds as long as the content does not have any identifiable information. This analysis also does not hold when using commercial services that identify their users to do billing. The only way to guarantee complete location privacy from access point owners in the subscription case is to use the occasionally connected computing model (which also defeats the purpose of subscribing to a WiFi service). It is likely that legal solutions and social conventions will be required, similar to those for cell phones. Note also that WiFi hotspots have a range of about 150 meters. So, depending on the configuration of access points owned by a single provider, there will be some inherent ambiguity and thus some room for some level of plausible deniability.

Access Point Privacy

Place Lab relies on the visibility of beacons generated by wireless Access Points (APs) for determining location. By re-purposing systems originally meant for wireless communications for location purposes, Place Lab makes AP owners stakeholders in the infrastructure even if they are not explicitly using the system. For example, AP owners might not want a system like Place Lab to divulge the exact location of their AP, nor perhaps the manufacturer or which security mechanisms are in use. One might argue that this information is already available from individuals who war drive and make this information public. However, new systems like Place Lab should not play to the lowest common denominator when it comes to privacy. If there is a high road with some "reasonable" level of cost then designers should take it. This high road is one way to address the co-evolving socio-technical concerns for location-aware systems.

However, it is difficult to protect information about APs in Place Lab directories because we expect people to store local copies of these databases. One possibility is to encrypt each entry in the database using each access point's MAC address as the key, but this scheme is hampered by the relatively small number of MAC addresses. MAC addresses are merely 48 bits in length, so any database that looked up information based only on MAC addresses would be vulnerable to an exhaustive search attack.[5] The situation is further complicated because in some cases, MAC addresses are manually set, and thus not globally unique.

To address these concerns, we can encode information about MAC addresses as pairs of adjacent or proximal pairs.[6] Thus, if in a given area, we have three APs with MAC addresses *P*, *Q*, and *R* near each other, where the

[5] The situation is somewhat more complicated than this. The first 2 bits are flags which are normally zero, the next 22 bits are traditionally the "organizationally unique identifier" (e.g., the manufacturer), and the remaining 24 bits are an organizationally unique address (e.g., an effective serial number for the given manufacturer). This acts to reduce the search space for an exhaustive search attack.

[6] What if there is only one AP in a given area, with no nearby or adjacent APs? If our database is only based on pairs of MAC addresses, this effectively makes these singleton AP useless for identifying location. Similarly, consider the case of a user who turns on her PDA and needs to discover her location immediately. If she is in range of two overlapping APs, she will be able to immediately discover her location, but otherwise, she will need to wait until she has moved far enough to come into range of a second AP to look up the information. We do not know if these are acceptable limitations, and experimentation is needed to discover what compromises are acceptable.

values are arranged so that $P < Q < R$, we can hash information about the APs using not a single MAC address but the pairs of MAC addresses *PQ* (i.e., *P* concatenated to *Q*), *PR*, and *QR*. In general, this will somewhat increase the complexity of mapping a given area for APs (adjacent APs will need to be recorded as such) and will slightly increase the number of entries we need to make in a hash table. Also, some experimentation will be necessary to find good compromises between the simplicity of a PDA or other mobile device and the size of the resulting database.

Once this infrastructure is in place, we can proceed immediately to protect the information using standard cryptographic mechanisms. Place Lab will use reliable one-way hash functions (denoted below as ***F***) to obfuscate details about APs such as location, manufacturer, and level of security unless a user can effectively demonstrate that they **could have** already known some of these details. This approach relies on the supposed uniqueness of a given pair of adjacent AP's MAC addresses. For example, suppose that *P* and *Q* are adjacent MAC addresses and that $P < Q$. Then a database look up on ***F*** (*PQ*) would yield

$$\textit{Database}(\boldsymbol{F}\,(PQ)) = \boldsymbol{E}_{PQ}(\textit{Position})$$

Where $\boldsymbol{E}_{PQ}$ denotes symmetric encryption using the key PQ. Assuming that pairs of MAC addresses are reasonably diverse (and thus not subject to an exhaustive search attack), and the ***E*** and ***F*** can satisfy normal security requirements (informally, that ***F*** can not be inverted, and that without *PQ*, ***E*** can not be inverted), it would not be possible for an adversary to determine a list of positions without knowing specific adjacent *PQ* values, which would presumably mean that the adversary had access to the physical location of the devices anyway. Therefore, this approach reveals no more information than would be available in any case — someone would need to actually travel to the location and observe which MAC addresses were adjacent. Note that end-users who are trying to calculate their locations are already able to see access points' MAC addresses.

It is important to point out that location-aware systems, such as Place Lab, can only go so far in protecting AP owners. As we have discussed above, Place Lab introduces no *new* vulnerabilities to the privacy of AP owners. AP owners who want protection must take action to secure their wireless networks. For the most basic security AP owners can use MAC based access control or enable WEP. More sophisticated access points can also be configured not to broadcast their MAC addresses. Although this makes it harder to discover wireless networks for legitimate users, it gives access point owners the choice of opting out.

Network Service Privacy

The model of client control over location data is appealing, but there are also potential attacks on the network. In many respects, this is the same problem faced by Mobile IP security. Fasbender et al have proposed using mixes that aggregate and redirect traffic, preventing the linking of senders and recipients and thus making it harder to do certain forms of traffic analysis [8].

Network service providers themselves would also have a difficult time identifying individual end-users, since most access points dynamically assign IP addresses.[7] Network service providers would be able to trace back to an access point owner, but not to specific individuals. Again, this analysis only holds if the individual does not subscribe to a WiFi service, and if the content transferred does not contain any personally identifiable information.

Web Service Privacy

Without some of the solutions in the network described above, web sites may well start correlating legitimately reported locations with IP addresses, eventually aggregating their own map of where a particular IP address is located, regardless of the user providing that information.

One workaround for individuals concerned about their privacy is to use services such as anonymizer.com that make web page requests on the user's behalf. Another possibility is to develop tools that periodically make fake requests to add chaff to the data.

It should also be noted that even if a web service can correlate IP addresses to physical locations, it may not be a significant threat to individual privacy if no personally identifiable information is transmitted. The web service might be able to infer that *someone* is there, but not necessarily who.

We expect that location usage policies will be worked into existing privacy policies and facilitated through mechanisms such as P3P. It is also possible that special server-side mechanisms will be developed to help manage location information. One idea is to let end-users download larger chunks of data at a time. For example, the Zagat's web site could be set up to let users download whole neighborhoods of information (zip code) at once rather than the restaurants within a single block (street). In the former case, the local device could then filter the chunked data to show just the information that is currently needed. This lets the end-user trade some level of convenience for privacy. That is, some more work is needed to filter the chunked data, but it makes it harder for others to know precisely where the user was or what she was looking for.

Another related idea is to have local devices pre-fetch information that *might* be useful but with some level of randomness for plausible deniability. For example, one could download not only the current neighborhood but two or three randomly selected nearby ones as well. Thus, pre-fetching could be used as a mechanism for protecting privacy, as well as improving availability and performance.

[7] Some people have suggested using open access points that do not require a password or a subscription to do file swapping because of the difficulty in tracking down individual users.

DISCUSSION

In this paper, we looked at some of the privacy issues in Place Lab, as well as architectural and user interface solutions for managing those issues. In this section, we pose some questions for discussion that look at how Place Lab fits into the larger picture of analyzing, designing, constructing, and evaluating privacy-sensitive applications for ubiquitous computing.

- The Place Bar represents one kind of privacy widget, a simple reusable component for managing privacy. What are other kinds of reusable widgets we can develop that provide feedback and control? It may be useful to have a similar tool that analyzes what location information a user has already disclosed to a web site and suggests what kind of tracking may be possible given the frequency and detail of the trail of locations divulged. Users could then set the level of disclosure at which they want to receive a warning. Work by Friedman, Howe, and Felten [9] has considered user interfaces for this type of informed consent for browser cookies and a similar model should be quite applicable for location information.
- The basic model of Place Lab is to have a decentralized architecture and start with data and services at the edge of the network. In other words, start with many personalized services for individuals rather than for groups to drive adoption. People might be more receptive to this approach because many ubicomp apps store personal data that some individuals simply do not feel comfortable sharing. Is this an approach that is likely to succeed, and can this approach be applied to other areas of ubicomp?
- Are there other simpler forms of ubicomp that we can be fostering as part of the larger goal of bootstrapping? As Mark Weiser noted, people often express their concerns about ubicomp in terms of privacy, when the underlying issue is often a lack of control and proper feedback about “what is controlling what, what is connected to what, where information is flowing, how it is being used” [10]. Using the adoption of the web as an analogy, a study by Pew Internet [11] noted that, initially, novices were often very concerned about privacy, but after a year of experience online, showed an increase in the number of trusting activities performed online. In other words, it is possible that “experience breeds higher levels of trust.” Simpler forms of ubicomp may help increase the levels of experience people have and could be an important factor in making people feel more in control.
- Privacy cuts across traditional layers in systems building. What kinds of hardware support, OS support, networking support, user interface support, and network services are needed to greatly simplify the task of creating privacy-sensitive apps? Also, what kinds of privacy mechanisms should be built in at each of these different layers?
- In addition to privacy, another goal of Place Lab is to provide libraries and APIs to make it easy to create location-enhanced web sites. However, designing privacy-sensitive apps is still an ad hoc process. What are better methods and tools for helping designers and developers understand or even predict what levels of privacy are needed? Can current methods such as contextual inquiry, participatory design, and design patterns, be adapted for privacy, and if so, how?
- What are the different concerns people have with respect to location privacy, and how important are they? For example, some dimensions of location privacy include spatial granularity (city → zip → street), temporal granularity (“I was at Tahoe sometime last month” rather than “I was at Tahoe July 1”), and temporal freshness (“You can have my location info if it is over a week old, but not my current location”). What are other dimensions, and what are their relative importance in different applications?

REFERENCES

1. Schilit, B.N., et al. *Challenge: Ubiquitous Location-Aware Computing*. in *The First ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots (WMASH '03)*. September 2003. San Diego, CA: ACM Press.
2. Schwartz, B., *The Social Psychology of Privacy.* American Journal of Sociology, 1968. **73**(6): p. 741-752.
3. Want, R., et al., *The PARCTAB Ubiquitous Computing Experiment*, in *Mobile Computing*, H.F. Korth, Editor. 1996.
4. Harter, A. and A. Hopper, *A Distributed Location System for the Active Office.* IEEE Network, 1994. **8**(1).
5. Ward, A., A. Jones, and A. Hopper, *A New Location Technique for the Active Office.* IEEE Personnel Communications, 1997. **4**(5): p. 42-47.
6. Priyantha, N.B., A. Chakraborty, and H. Balakrishnan. *The Cricket Location-Support System*. in *MobiCom 2000: The Sixth Annual International Conference on Mobile Computing and Networking*. 2000. Boston, Massachusetts: ACM Press.
7. Bahl, P. and V.N. Padmanabhan. *RADAR: An In-Building RF-Based User Location and Tracking System*. in *IEEE INFOCOM 2000*. 2000. Tel-Aviv, Israel.
8. Fasbender, A., D. Kesdogan, and O. Kubitz. *Analysis of Security and Privacy in Mobile IP*. in *4th International Conference on Telecommunication Systems, Modeling and Analysis*. 1996. Nashville, TN.
9. Friedman, B., D.C. Howe, and E. Felten. *Informed Consent in the Mozilla Browser: Implementing Value-Sensitive Design*. in *The Thirty-Fifth Annual Hawai'i International Conference on System Sciences*. 2002: IEEE Computer Society.
10. Weiser, M., R. Gold, and J.S. Brown, *The Origins of Ubiquitous Computing Research at PARC in the Late 1980s.* IBM Systems Journal, 1999. **38**(4): p. 693-696.
11. Pew Internet & American Life, *Testimony of Lee Rainie: Director, Pew Internet & American Life Project.* 2001.