Hitchhiking: A Privacy-Preserving Framework for Collecting Location-Based Data on Commodity Devices

Karen P. Tang[1], James Fogarty[1], Pedram Keyani[1], Anket Mathur[2], Jason I. Hong[1]
[1]Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213
{kptang, jfogarty, pkeyani, jasonh}@cs.cmu.edu

[2]Computer Science Department
University of Texas - San Antonio
San Antonio, TX, 78249
amathur@cs.utsa.edu

ABSTRACT
The emergence of location-based computing promises new and compelling applications, but raises very real privacy risks. Existing approaches to privacy generally treat people as the entity of interest, often using a fidelity tradeoff to manage the costs and benefits of revealing a person's location. However, these approaches cannot be applied in some applications, as a reduction in precision can render location information useless. This is true of a category of applications that use location data collected from multiple people. We present hitchhiking, a new approach that treats locations as the primary entity of interest. Hitchhiking removes the fidelity tradeoff by preserving the anonymity of reports without reducing the precision of location disclosures. We can therefore support the full functionality of an interesting class of location-based applications without introducing the privacy concerns that would otherwise arise.

1. INTRODUCTION
The pending ubiquity of location-based applications has significant implications for anonymity and privacy. Consider an otherwise anonymous person who starts almost every day in a given location and ends the day in that same location. An application that is able to collect this data can identify the person by checking a database to see who lives at that address. There is also a potential for individuals to abuse location-based applications for more malicious purposes, targeting a specific victim and obtaining his location and movement information.

Significant prior work has examined anonymity and privacy in location-based applications [1,2,3,4,5], but generally make two assumptions: *first, prior work generally treats a person as the entity of interest and second prior work often treats location privacy as a fidelity tradeoff.* For example, a person might reveal their location as part of a query about their surroundings or as a part of a social interaction with friends. Revealing a more precise indication of one's identity or location often implies better services can be offered. This has led prior work to focus on techniques for balancing the fidelity of disclosure against the utility of an application.

As an alternative solution, we present *hitchhiking*, a new approach to anonymous and privacy-sensitive collection of sensed data in location-based applications. *Hitchhiking applications treat locations as the entity of interest.* Because the knowledge of who is in a location is irrelevant, the fidelity tradeoff is removed. Instead, *hitchhiking ensures the anonymity of people providing information about a location.* We can therefore obtain the full functionality of an interesting class of location-based applications without the privacy concerns that would otherwise arise.

2. HITCHHIKING APPROACH
Our work can be considered an example of a privacy risk model (as shown in [3]): we have identified the privacy threats encountered in a category of location-based applications and have developed strategies for addressing these threats. We consider a person's anonymity or privacy to have been violated in either of two scenarios: 1) An *identity violation* has occurred if a single report reveals a person's identity, or 2) A *tracking violation* has occurred if a report can be identified as being provided by the same person who provided an earlier report. Figure 1 lists four categories of threats that can result in identity or tracking violations. The hitchhiking approach addresses these threats with seven requirements (Figure 2).

1)	Collected location logs can be abused by a server operator or by other people who gain access.
2)	A user could be targeted by monitoring their home or another similarly sensitive location.
3)	A location approval could be spoofed, tricking a target user into approving a sensitive location.
4)	By hiding an identifier in a location definition, a server could track when people visit a location.

Figure 1. Four categories of potential location-based privacy threats that could occur in hitchhiking applications.

1)	Location is computed on the client.
2)	Only the client device is trusted.
3)	Each person approves reporting from a location.
4)	Physical constraints prevent location spoofing.
5)	Location identifiers are based on physical location.
6)	Location identifiers are generated by the client.
7)	Sensed identifiers are not reported to the server.

Figure 2. Seven requirements of hitchhiking that protect users' privacy in location-based applications.

These seven requirements combine to ensure that a malicious server cannot induce identity or tracking violations. Each person approves every location they report from, and the use of physical constraints ensures that a spoof cannot mask what location the person is approving. Because none of the information in a report

was initially provided by the server, there is no opportunity for the server to hide an identifier in the report. The server knows the physical properties of each location (such as the GPS coordinates of a highway or the WiFi access points in a coffee shop), so it can infer what location is being reported on. But the server cannot infer who made a report. Hitchhiking is therefore able to preserve privacy by anonymizing the user and effectively decoupling the location-based data from the user who provided the data. An indirect benefit from hitchhiking also arises from the observation that by removing the link between the data and the user, hitchhiking has also simplified the data hierarchy on the server, making it much easier for system administration to diagnose any potential security problems that may arise on the server. It is also worth addressing that, while location-based applications can be implemented in different ways, hitchhiking warrants consideration because it requires no additional infrastructure. Based entirely in software on devices that people already carry, hitchhiking applications can be deployed at extremely low cost while still preserving the user's privacy for location-based applications.

3. APPLYING HITCHHIKING

Hitchhiking supports a general category of applications that collect sensed data from locations of interest. In this section, we apply Hitchhiking to an application as a demonstration of the breadth of our approach. The goal of hitchhiking is to preserve the full desired functionality of these applications while removing privacy threats that would otherwise arise.

3.1 Bustle: Monitor Coffee Shop Availability

Bustle is a location-centric service that senses WiFi-networked laptops and anonymously reports estimates of table availability in coffee shops. In a typical usage scenario, a person visits a coffee shop and works on his laptop. Running in the background, Bustle scans for nearby WiFi access points to see if the person is in a coffee shop [4]. After determining it's okay to report, Bustle monitors Address Resolution Protocol (ARP) broadcasts to determine how many other devices are present. At regular intervals the laptop reports the counts to a server, which then infers from these counts the shop's busyness.

We conducted a small feasibility study of Bustle to test if the correlation between laptop usage and the number of people in a coffee shop is sufficient for inferring space availability. We made 20 visits to a laptop friendly coffee shop over 7 days, spacing visits by at least 90 min, aiming for 9AM-9PM coverage. On each visit, we monitored ARP broadcasts for 20 minutes and then counted the empty tables. In the shop we sampled, there is a strong correlation between the number of computers on the network and the number of empty tables. In every case of no available tables, at least 8 computers were detected on the network. While the strength of this correlation will vary in different places, this result shows a learnable threshold for Bustle's WiFi-based busyness sensing.

3.2 Other Hitchhiking Domains

The location-based privacy threats (Figure 1) and the resulting hitchhiking counters (Figure 2) can apply to other location technologies and to other location-centric applications as well. Examples of such applications include monitoring traffic, tracking bus routes, and monitoring conference room availability [6].

4. CONCLUSION

We have presented hitchhiking, an anonymous and privacy-sensitive approach to a category of location-based applications. The fundamental tenet of hitchhiking is that reports are always strictly about a location and cannot be tied to a person. By presenting a privacy risk analysis of hitchhiking, this paper provides designers of location-based applications and services with an approach to building a useful class of application while also protecting end-user privacy. Implemented entirely in software on the client device, hitchhiking does not require new hardware or a trusted middleware platform. It is therefore possible to deploy applications on existing phone and WiFi networks, without the active cooperation of the network provider. By enabling anonymous and privacy-sensitive data collection, hitchhiking protects users and removes personal privacy as an obstacle to a category of location-based applications.

5. ACKNOWLEDGMENTS

We thank all of the contributors to Place Lab, jpcap, libpcap, and the JDesktop Integration Components. This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. NBCHD030010, by an AT&T Labs fellowship, and by the National Science Foundation under grants IIS-0121560 and IIS-032531.

6. REFERENCES

[1] Beresford, A.R. and Stajano, F. Location Privacy in Pervasive Computing. *IEEE Pervasive Computing, 2*(1).

[2] Gruteser, M. and Grunwald, D. Anonymous Use of Location-Based Services Through Spatial and Temporal Cloaking. *Proceedings of the ACM Conference on Mobile Systems, Applications, and Services (MobiSys 2003)*, 31-42.

[3] Hong, J.I., Ng, J.D., Lederer, S. and Landay, J. Privacy Risk Models for Designing Privacy-Sensitive Ubiquitous Computing Systems. *Proceedings of the ACM Conference on Designing Interactive Systems (DIS 2004)*, 91-100.

[4] Schilit, B.N., LaMarca, A., Borriello, G., Griswold, W.G., McDonald, D., Lazowska, E., Balachandran, A., Hong, J.I. and Iverson, V. Challenge: Ubiquitous Location-Aware Computing and the Place Lab Initiative. *Proceedings of the ACM International Workshop on Wireless Mobile Applications and Services on WLAN (WMASH 2003)*, 29-35.

[5] Smith, I., Consolvo, S., Hightower, J., Iachello, G., LaMarca, A., Scott, J., Sohn, T. and Abowd, G. Social Disclosure of Place: From Location Technology to Communications Practices. *Proceedings of the International Conference on Pervasive Computing (Pervasive 2005)*, 134-151.

[6] Tang, K.P., Keyani, P., Fogarty, J., Hong, J.I. Putting People in their Place: An Anonymous and Privacy-Sensitive Approach to Collecting Sensed Data in Location-Based Applications. *Proceeding of the ACM Conference on Human Factors in Computing Systems (CHI 2006).*