Moving forward: Location privacy and location awareness

Matt Duckham
University of Melbourne
Victoria 3010, Australia
mduckham@unimelb.edu.au

ABSTRACT

This extended abstract looks forward at which shared concepts and approaches are shaping future research in the field of location privacy, as well as reflecting on the broader achievements and progress that the field can claim. Seven key principles of research into location privacy are proposed. Taken together, these principles aim to delineate what makes research into location privacy and location-awareness different and distinctive, when compared with other research topics in privacy and spatial information science more generally.

Categories and Subject Descriptors

C.2.4 [**Computer-communication networks**]: Distributed systems; K.4.1 [**Computers and society**]: Public policy issues—*Privacy*

General Terms

Theory, Legal aspects

Keywords

location-based services, location privacy, spatial information, decentralization, movement patterns, user-generated content, accuracy, precision

1. INTRODUCTION

Location privacy concerns the right of individuals to control the collection, use, and communication of personal information about their location. As such, location privacy is a special case of *information privacy*, the "claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" [1]. Location-aware computing concerns the automatic use of knowledge about an individual's location in order to provide more relevant information or information services to that individual. Location-aware computing is similarly a special case of context aware computing [2]. Information systems and services that rely on location-aware computing are typically called *location-based services*.

SPRINGL '10 November 2, 2010. San Jose, CA, USA

The fundamental challenge facing research into computing with private location data, is to balance these two competing aims of location privacy and location awareness: providing better access to information related to a person's location, while still enabling that person to retain control over the spread of knowledge about their location. This extended abstract proposes seven key principles that can be identified from location privacy research over recent years. While these principles represent a personal view of what makes location privacy research special, they are chosen to promote reflection on what has been achieved in the past, and discussion about how best to consolidate and progress into the future.

Principle #1: Geographic space presents constraints to movement

Except—perhaps—for users of location-based services sailing on the high seas, piloting an aircraft, or trekking in the world's great deserts, geographic space *always* presents constraints to movement. Location-based service users must always negotiate buildings, roads, borders, lanes, topography, and a myriad other constraints to movement. Thus an important principle, not always adhered to in past research on location privacy, is that location privacy protection must assume constraints to the movement of location-based service users. For example, privacy research usually begins with the default assumption of movement through a network space, rather than assuming users can move freely on the plane. Failing to account for the constraints to movement presented by geographic space is a problem because privacy adversaries will surely account for these constraints. Thus, if a privacy protection approach assumes unconstrained movement within the plane, it will be vulnerable to attacks where an adversary uses knowledge of the constraints to movement (like a road network, building layout, topographic map) to refine its knowledge about the location of a location-based service user moving amongst those constraints.

Principle #2: Humans are not random

A common feature of early work into location privacy was to begin with the assumption that humans are random. Examples of assumptions in this category include that humans are randomly distributed through space for privacy queries, or that humans follow (correlated or constrained) random walks when moving through space. Such assumptions make modeling and simulating human movement practically and mathematically convenient. However, in the real world humans are never randomly distributed, never move randomly, never query at random time intervals, and indeed rarely engage in any purposeful activities that can be usefully regarded as random processes. While randomness can be important in the design of experiments, assuming randomness again increases a location privacy protection strategy's vulnerability to adversarial attack.

In these attacks, an adversary with knowledge of the non-random structure of human locations can subvert privacy protection, and refine its knowledge of a person's location.

Principle #3: Large user-contributed data sets are biased

Principles #1 and #2 have important implications for the design of location privacy protection strategies, as well as for the design of experiments to evaluate the performance of privacy protection strategies. It is hard to simulate realistic human movements, because they are not random and not unconstrained. Thus, as more user generated content (UGC) becomes available, it is tempting to turn instead to large user-contributed data sets to evaluate location privacy protection procedures. However, here lies another potential problem. The Pareto principle has particular relevance to such data sets, because it is to be expected that a large proportion of the data will be contributed by a small proportion of users (the so-called "80-20 rule," typically 80% of data comes from 20% of users). Such features are well known in studies of UGC (e.g., [3]), and are also highly relevant to location privacy studies. For example, it is expected that user contributed data sets of movements will not be entirely representative of human movement more generally. Consequently, evaluating privacy protection techniques using UGC needs to be handled carefully to ensure that any apparent benefits have broader significance, and are not simply a feature of the movements of the minority of people who contribute the majority of UGC.

Principle #4: Continuous and snapshot queries are different

Privacy protection for snapshot queries is now a well-studied topic. Approaches to snapshot queries are easy to define and explore because they are essentially *stateless*: all the knowledge required for the query (and so available to a privacy adversary) is contained in the query. For this reason, privacy protection for continuous, long running queries is harder to provide—adversaries can build up knowledge of a user's location over time. However, adversaries monitoring continuous queries can also refine their knowledge of an individual's location by examining the spatial and temporal relationships embedded in information about movement over time. For example, by making assumptions about the maximum speed of movement of a person can enable an adversary to rapidly refine knowledge of a person's location even when each snapshot may be protected using techniques spatial cloaking, dummies, or obfuscation [4]. Thus, privacy protection for continuous queries requires techniques that go further than simple, repeated applications of a snapshot query privacy protection technique, and as a result is a important and substantially more challenging topic.

Principle #5: Location privacy attacks are as important as location privacy protection

If information is worth protecting, it is also worth attacking. An important principle of today's location privacy research, not always evident in earlier research, is the central role of an analysis of the counter strategies and threats to location privacy protection. These two topics, privacy protection and privacy attacks, are simply two sides of the same coin; one cannot exist without the other. Thus, part of the evaluation of any privacy protection strategy must always be an analysis of the threats to privacy protection.

Principle #6: Decentralization does not always improve location privacy

A promising and rapidly emerging approach to location privacy protection is to take advantage of decentralized, peer-to-peer computing architectures for location privacy protection. Such approaches typically recruit nearby peers to assist in protecting or even answering individual queries, minimizing the information that is communicated to and stored in centralized servers. These approaches are surely less vulnerable to some types of privacy attacks, such as intrusive inferences, where private information about a user like a home address is inferred from patterns of movement. An adversary hoping to make intrusive inferences will be more likely to be confounded if there exists no single, centralized repository of sensitive location data. However, decentralization may be *more* vulnerable to other types of privacy attack, in particular physical harm (like stalking). An adversary threatening physical harm is likely to already need to be in close spatial and temporal proximity to an intended target. Thus, for these types of attack there is good reason to expect that adversaries may exist amongst the spatially and temporally nearby peers—the same peers amongst whom sensitive information is typically shared in decentralized location privacy protection approaches.

Principle #7: Accuracy and precision are not synonyms

Although the fact that accuracy and precision are not synonymous can hardly be considered a "principle" of privacy research, it is surprising how often in the past fundamental terms like accuracy and precision are misused. Accuracy concerns the level of correctness in information. Precision concerns the level of detail in information. Accuracy and precision are entirely orthogonal: information can be accurate but not precise and vice versa, as well as both/neither accurate and/nor precise. The distinction between accuracy and precision is especially relevant to a class of location privacy protection strategies that attempt to hide location information (like spatial/temporal cloaking, obfuscation, and location dummies [5, 6, 7]). Both accuracy and precision can be used to hide information about a person's location. Imprecision, for example, is the basis of temporal cloaking, where a user may decrease the level of temporal detail about their location that is provided to a location-based service provider. Similarly, location dummy strategies typically use imprecision, by communicating a user's true location along with a number of other false locations (dummies). Although each individual dummy provides *inaccurate* information about the user's location (i.e., it does not correctly represent where the user is located), the set of dummy locations together with the user's true location constitutes *imprecise* information about the user's location (i.e., it provides limited detail about at which of the set of locations the user is actually located). Only if the set of dummy locations does not contain the user's true location can it be considered inaccurate (as well as imprecise). Distinguishing correctly between these fundamental concepts, accuracy and precision, may seem a minor issue, but such issues are prerequisites for establishing and maintaining rigor and credibility for the field.

2. CONCLUSIONS

While the seven principles chosen in this paper represent a personal view, it is not difficult to find concrete examples in the literature of articles do, and decreasingly do not, exhibit these principles. Others may reasonable argue for different choices. Overall, however, there are some clear patterns in what makes research into computing with location privacy special. First, geographic space is highly structured and correlated, not random, and so adversaries

may take advantage of this structure to subvert location privacy protection (principles #1, 2, and 4). Second, location privacy protection operates within the milieu of location-aware computing, and as a result must accommodate the associated spatial constraints to computing with location information (principles #3 and #6). Third, it is important that location privacy research continue to situate itself within the context of of research into privacy and spatial information science more broadly, adopting and building on principles and concepts already established in those fields (principles #5 and #7).

Acknowledgements

Matt Duckham's research is supported under the Australian Research Council's Future Fellowship scheme (project number FT0990531).

3. REFERENCES

[1] A. F. Westin, *Privacy and freedom*. New York: Atheneum, 1967.

[2] A. Schmidt, M. Beigl, and H.-W. Gellerson, "There is more to context than location," *Computer and Graphics Journal*, vol. 23, no. 6, pp. 893–902, 1999.

[3] L. Hollenstein and R. Purves, "Exploring place through user-generated content: Using flickr tags to describe city cores," *Journal of Spatial Information Science*, no. 1, pp. 21–48, 2010.

[4] M. Duckham, L. Kulik, and A. Birtley, "A spatiotemporal model of obfuscation strategies and counter strategies for location privacy," in *Geographic Information Science* (M. Raubal, H. Miller, A. Frank, and M. Goodchild, eds.), vol. 4197 of *Lecture Notes in Computer Science*, pp. 47–64, Berlin: Springer, 2006.

[5] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proc. MobiSys '03*, pp. 31–42, 2003.

[6] M. Duckham and L. Kulik, "A formal model of obfuscation and negotiation for location privacy," in *Pervasive 2005* (H. W. Gellersen, R. Want, and A. Schmidt, eds.), vol. 3468 of *Lecture Notes in Computer Science*, pp. 152–170, Berlin: Springer, 2005.

[7] H. Kido, Y. Yanagisawa, and T. Satoh, "Protection of location privacy using dummies for location-based services," in *Proc. 21st Data Engineering Workshop*, p. 1248, IEEE, 2005.