Abstract—The popularity of location-based services leads to serious concerns on user privacy. A common mechanism to protect users’ location and query privacy is spatial generalisation. As more user information becomes available with the fast growth of Internet applications, e.g., social networks, attackers have the ability to construct users’ personal profiles. This gives rise to new challenges and reconsideration of the existing privacy metrics, such as k-anonymity. In this paper, we propose new metrics to measure users’ query privacy taking into account user profiles. Furthermore, we design spatial generalisation algorithms to compute regions satisfying users’ privacy requirements expressed in these metrics. By experimental results, our metrics and algorithms are shown to be effective and efficient for practical usage. General Terms: Security, measurement

Keywords
Location based services, query privacy, anonymity, measurement

I. INTRODUCTION

Combined technological advances in location-sensing, mobile computing, and wire-less communication are opening up new and exciting opportunities in the domain of location-aware computing. Many of these opportunities are explored elsewhere in this others are already being developed into practical applications that will provide benefit to a wide cross-section of society, such as elder care [6], emergency response and E911 systems [7], and navigation for the visually impaired [3].

Despite the undoubtedly future potential of location-aware computing, location-awareness also presents inherent future threats, perhaps the most important of which is location privacy. Most people would not feel comfortable if regularly updated in-formation about their current location were made public, any more than we would feel comfortable if information about our home address, telephone number, age, or medical history were public. Our precise location uniquely identifies us, more so than our names or even our genetic profile. Privacy is regarded as a fundamental human right, internationally recognized in Article 12 of the UN Universal Declaration of Human Rights. The history and development of privacy rights have been examined from many different perspectives in the literature [4] for a concise overview of the history of privacy from the perspective of ubiquitous and location-aware computing).

Not all authors agree that privacy should be regarded as an inalienable right. Some authors have argued for greater transparency in place of privacy. Proponents of greater transparency cite the practical difficulties of protecting privacy in the face of changing technological capabilities (encapsulated in the now infamous remark by Sun CEO Scott McNealy: “You have zero privacy anyway, get over and the public benefits that may be accrued through the relaxation of some privacy.

Studies of users’ attitudes to location privacy issues often provide some support for these views. Evidence presented in indicates a lack of awareness or even moderate indifference to location privacy issues amongst the general public. Other studies have painted a more complex picture. For example, Barkuus and Dey [5] found that concern about location privacy can be dependent on the type of application, with applications that track users’ movements over a period of time causing more concern than simple positioning applications.

Attitudes to privacy have changed in the past and will continue to change over time. Although the need for a right to privacy will continue to be debated, in the shorter term at least there would seem to be a pressing need for privacy protection measures able to cope with a rapidly changing technological landscape. Concerns about protecting the individual’s right to privacy have previously appeared in connection with numerous other new technologies, including GIS [4], the Internet [1], and collaborative user interfaces [3]. The need for location privacy is recognized in some of the earliest literature on information privacy (e.g., [7]) and location-aware computing. Looking at more recent literature, it is possible to identify at least three key negative effects associated with failures to protect location privacy within a location-aware computing environment.

II. POSITIONING SYSTEMS AND LOCATION PRIVACY

In addition to the social constraints on location privacy,
discussed in the previous section, location-aware computing environments place certain technical constraints on location privacy. The primary technical constraints arise from the positioning systems themselves. Hightower and Boriello provide a survey of the wide variety of positioning systems currently in use [3]. In addition to the familiar GPS, positioning systems in the literature and in common usage include triangulation of RF wireless LAN signals (e.g., [4]), proximity to infrared beacons (e.g., [6]), scene analysis and computer vision (e.g., [4]), and inertial tracking. New positioning systems, such as audio-based positioning [6] and radio signal profiles [4], are continually being developed.

There exist several classifications of positioning systems. For example, a top-level distinction is often made between active positioning systems, which rely on the establishment of beacons to operate (such as WiFi signal triangulation, GPS, infrared proximity sensors), and passive positioning systems, which require no beacons (such as inertial navigation, scene analysis, and audio-based positioning, see [7] for more information). However, from a privacy perspective, positioning systems are more usefully classified into client-based, network-based, and network-assisted systems [5].

- In client-based positioning systems, mobile clients autonomously compute their own location (for example GPS and inertial navigation). It is technically possible in a client-based positioning system for a client to compute its location, without ever revealing that location to any other entity.

- In network-based positioning systems, the network infrastructure is responsible for computing a mobile client's location. Cell phone phone positioning using CGI (cell global identity) is an example of network-based positioning. In network-based positioning systems, the network infrastructure administrator must hold information about the location of mobile clients.

- In network-assisted positioning systems, a combination of client-based and network-based computation is required to derive a client's location. For example, A-GPS (assisted GPS) combines network-based CGI positioning to increase the speed of GPS positioning. In network-assisted positioning systems, some information about a mobile client's location must reside in the network infrastructure, although this information may be less precise than the information held by the mobile client itself.

Downloading spatial data sets from a remote service provider will be subject to wireless network bandwidth limitations and may provide an indication of the user's location (either by inferring location from knowledge of the data sets of interest to the user or by positioning using a client's mobile IP address, as in [1]). Alternatively, storing all potentially useful spatial data in a user's mobile device leads to the data integrity and currency issues that are inevitably associated with maintaining copies of the same data sets across multiple clients.

### III. LOCATION PRIVACY PROTECTION STRATEGIES

Having identified location privacy as a key issue for location-aware computing and outlined some of the technical aspects of location privacy, the next step is to ask what mechanisms exist for location privacy protection. The different strategies that exist for protecting a mobile individual's location privacy can be classified into four categories: regulatory, privacy policies, anonymity, and obfuscation strategies. In this section each type of strategy is reviewed in turn.

#### 3.1 Regulatory strategies

Regulatory approaches to privacy involve the development of rules to govern fair use of personal information. Most privacy regulation can be summarized by the five principles of *fair information practices* (originally developed as the basis of the US privacy legislation)

1. **Notice and transparency:** Individuals must be aware of who is collecting personal information about them and for what purpose.

2. **Consent and use limitation:** Individuals must consent to personal information being collected for particular purposes, and the use of personal information is limited to those purposes.

3. **Access and participation:** Individuals must be able to access stored personal data which refers to them, and may require that any errors be corrected.

4. **Integrity and security:** Collectors must ensure personal data is accurate and up-to-date and protect against unauthorized access, disclosure, or use.

**Enforcement and accountability:** Collectors must be
accountable for any failures to comply with the other principles.

Although these principles of fair information practice are at the core of most privacy regulation (e.g., [50, 66]), there are a variety of ways in which these rules have been implemented. In general, regulatory frameworks aim to adequately guarantee privacy protection for individuals without stifling enterprise and technology. The concept of co-regulation, which aims to encourage flexible self-regulation on top of legal enforcement of minimum privacy standards, is one example of a mechanism for achieving such a balance [1].

The concept of fair information practices is usually applied to “personal information” in general, not specifically to location information. Personal information can be defined as “information … about an individual who's identity is apparent, or can reasonably be ascertained, from the information [3].” In this respect, location in formation is usually treated as one type of personal information, like age, gender, or address. A small number of privacy regulations have been developed to address location privacy issues explicitly (for example, proposed location tracking legislation in Korea [5] and the discontinued AT&T “Find Friends” location-based service [6]).

Although regulation lies at the foundations of any privacy protection system, there are at least four reasons for believing that, on their own, regulations do not represent a complete solution to location-privacy concerns. First, regulation itself does not prevent invasions of privacy. It simply ensures that there exist mechanisms for “enforcement and accountability” when unfair information practices are detected. Second, the development of regulation may lag behind innovation and new technology. Third, regulation applies “across the board,” making a satisfactory balance between guaranteed levels of privacy protection and freedom to innovate and develop new technology difficult to achieve, even using models such as co-regulation. As a consequence, other privacy protection mechanisms are needed in addition to regulation. Finally, abiding by fair information practice principles can give rise to practical problems with respect to location-awareness. For example, Ackerman et al. [2] examine the difficulties created by the requirements for notice and consent for user interfaces and HCI in context-aware computing environments.

3.2 Privacy policies

Privacy policies are trust-based mechanisms for proscribing certain uses of location information. Whereas regulation aims to provide global or group-based guarantees of privacy, privacy policies aim to provide privacy protection that is flexible enough to be adapted to the requirements of individual users and even individual situations and transactions. Overviews of a range of different privacy policy systems can be found in [2]. In this section we summarize three of the major privacy policy initiatives currently underway that illustrate the range of approaches that privacy policies can take.

3.3 IETF GeoPriv

The Internet Engineering Task Force (IETF) is an international consortium concerned with future Internet architectures. The IETF’s GeoPriv working group is adapting PIDF (presence information data format) as a privacy policy system for location privacy. PIDF is an IETF XML dialect for instant messaging, which includes a mechanism for exchanging information about the presence of a person (or place or thing) [5]. The GeoPriv specification additionally includes information about the location of that person, effectively annotating location data with metadata about the fair uses of that location data. In order to protect location privacy, the GeoPriv specification defines a location object which encapsulates both an individual's location and their privacy policy. At the center of the privacy policy are usage rules which describe acceptable usage of the information, such as whether retransmission of the data is allowed or at what date the information expires, and must be discarded. Further, location objects can be digitally signed, making the privacy policy resistant to separation from the location information [4].

3.4 PDRM

Digital rights management (DRM) concerns the technical efforts by some intellectual property vendors and other organizations to enforce intellectual property protection (for example, protection from piracy). PDRM (personal DRM) adopts a similar approach for personal data. When applied to location privacy, the PDRM approach is closer to the “user-oriented” IETF GeoPriv model than the P3P “provider-oriented” model. For location-aware systems, location data is treated as the property of the person to whom that data refers. PDRM then aims to enable that person to “license” the personal data for use by a location-based service provider [29]. So, for example, an entity wishing to use an individual’s location data may first need to demonstrate their willingness to agree to the licensing, which may set limits on that entity’s ability to share or process the Policy-based initiatives for privacy protection, like PDRM, P3P, and GeoPriv, are continuing to develop. However, there are again reasons for believing that policy-based initiatives provide only a partial answer to the question of location privacy protection. First, privacy policies are often highly complex and their practicality for use in location-aware environments with frequently updated highly dynamic information remains, as yet, unproven. Second, privacy policies systems generally cannot enforce
privacy, instead relying on economic, social, and regulatory pressures to ensure privacy policies are adhered to. Consequently, privacy policies are ultimately vulnerable to inadvertent or malicious disclosure of personal information.

### 3.5 Anonymity

Anonymity concerns the dissociation of information about an individual, such as location, from that individual's actual identity. A special type of anonymity is pseudonymity, where an individual is anonymous, but maintains a persistent identity (a pseudonym) [53]. For example, [20] describe a location-aware system for allowing users to leave and read digital notes at specific locations (“geonotes”). One of the ways users can protect their privacy is to associate an alias (pseudonym) with a note in place of their real name.

An explicitly spatial approach to providing anonymity in location-aware computing environments is presented in [27]. Gruteser and Grunwald used a quad-tree-based data structure to examine the effects of adapting the spatial precision of information about an individual's location according to the number of other individuals within the same quadrant, termed “spatial cloaking.” Individuals are defined as $k$-anonymous if their location information is sufficiently imprecise in order to make them indistinguishable from at least $k - 1$ other individuals. The authors also explore the orthogonal process of reducing the frequency of temporal information, termed “temporal cloaking.”

### 3.6 Zero-knowledge proofs

The idea of a zero-knowledge proof is to prove the knowledge of a certain fact without actually revealing this fact. Zero-knowledge proofs (ZKPs) involve a prover, who attempts to prove a fact, and a verifier, who validates the prover's proof. The verifier may determine the correctness of the proof, but does not learn how to prove the fact or anything about the fact itself. Fiat and Shamir developed the first practical zero-knowledge proof system in 1987 [23].

ZKPs often appear somewhat counter-intuitive at first, so consider the following simple example. Person $A$ claims to know the secret combination to a safe. Person $B$ deposits a valuable item in the safe, locks the safe, and leaves the room without the safe. Person $B$ does not know the combination to the safe. If person $A$ is able to present the item locked in the safe to $B$, then $A$ has proven to $B$ that $A$ knows the combination to the safe without revealing the actual combination. In ZKP terminology, the proof is interactive because the verifier (person $B$) challenged the prover (person $A$) and the prover must respond to the verifier.

In a ZKP, a prover may provide the correct response to a challenge purely by chance. To combat this possibility, there are usually several rounds of challenges and responses in a ZKP. As the number of rounds increases, the probability that the prover will give the correct answer in every round decreases. Typical ZKPs will verify a proof with a probability of $1 - 1/2^n$, where $n$ is proportional to the number of rounds used.

There are two distinct application scenarios for ZKPs:

- **Authentication**: Prover $P$ is able to prove to verifier $V$ that $P$ is authorized to access information without requiring any knowledge about $P$'s identity.

- **Identification**: Prover $P$ can prove to verifier $V$ that $P$ is $P$, but no party $Q$ is able to prove to $V$ that $Q$ is $P$.

The first application scenario that uses ZKPs without revealing an individual's identity is **anonymous digital cash** [9]. To date, ZKPs have not been widely re-searched within the domain of location-aware computing. However, clearly ZKP-based authentication and identification might also be used with location-based services, and initial work in this area is beginning to appear (e.g., [1]).

There is one further, explicitly spatial problem facing any anonymity-based system for location privacy: a person's identity can often be inferred from his or her location. Consequently, anonymity strategies (even those employing pseudonymity or ZKPs) are vulnerable to data mining [19]. Berserfor and Stajano [7] have used simulated historical data about anonymized individual's movements to investigate ways of subverting anonymity-based privacy protection. Their results show how simple heuristics can be used to deanonymize pseudonyms, providing users with much lower levels of location privacy than might naively expected. Thus, anonymity alone cannot hope to provide total location privacy protection.

### 3.7 Obfuscation

Obfuscation is the process of degrading the quality of information about a person's location, with the aim of protecting that person's location privacy. The term “obfuscation” is introduced in [1], but several closely related concepts have been proposed in previous work. The “need-to-know principle” aims to ensure that individuals release only enough information that a service provider needs to know in order to provide the required service [3]. The idea of a need-to-know principle is closely related both to obfuscation and the fundamental fair information practice principle of consent and use. Snekennes investigates a privacy policy-based approach to enforcing the need-to-know principle in location-aware computing by adjusting precision of location information [6]. In the domain of anonymity-based approaches, the work of Gruteser and Grunwald aims to enforce the “principle of minimal collection” [2], again akin to obfuscation. On a slightly different theme, Jiang et al. discuss the “principle of minimal asymmetry,” which aims to ensure that the flow of personal information away from an individual is more...
closely matched by the information flow back to that individual about who is using that information for what purposes.

IV. CONCLUSION

Location privacy lies at the intersection of society and technology. This chapter has reviewed the reasons why location privacy is becoming such an important topic in society, and the technological constraints to location privacy. When considering the strategies that can be used to protect an individual’s location privacy, it becomes clear that no single strategy currently available is capable of providing a complete solution to location privacy protection. Each approach has distinct advantages and dis-advantages. Therefore, it seems likely that the future of location privacy protection involves combinations of the approaches: regulation, privacy policies, anonymity, and obfuscation.

There remain many challenges for privacy researchers. For example, for information to be worth protecting, it must also be worth attacking. Current research tends to be biased toward privacy protection. By contrast, it is also important to understand the techniques a hostile agent might employ in order to invade a person's privacy (circumventing location privacy protection and attempting to discover an individual's exact location). In this respect, privacy research is analogous to cryptology, which comprises both cryptography (code making) and cryptanalysis (code breaking).

As this chapter has shown, location information differs from many other types of personal information. Consequently, future research aimed specifically at location privacy will need to focus on specialized privacy protection techniques for several reasons. First, unlike many other types of personal information, identity may be inferred from location. Second, information about personal location is highly dynamic. By contrast, current research approaches to location privacy are usually fundamentally static in nature, modeling the movement of an individual as a sequence of static snapshot locations. Many aspects of location privacy demand models that provide a more faithful representation of the temporal aspects of LBS. Finally, the potential uses of spatial information are highly varied. Correspondingly, the potential benefits of invading an individual’s location privacy may be higher than for some other types of information. Without proper protection, the location information generated by location-aware systems could conceivably be abused or unfairly used in almost any domain of human, social, or economic activity, including marketing, insurance, surveillance, harassment, social security, politics, law enforcement, health, or employment. Indeed, it is this very feature of location information that makes location information so vital to our future information systems.

REFERENCES


