Location Privacy: User-centric Threat Analysis

Benjamin Greschbach
School of Computer Science and Communication
KTH Royal Institute of Technology
Stockholm, Sweden
bgre@kth.se

Information that describes the geographic locations of a person over time is a fairly new class of potentially privacy-harming data. In pace with certain technological advances of the recent years, more and more location data is generated and processed by various systems. Its usage for different location-based services (including the integration into social network services) encounters a steep and still ongoing rise in popularity. Besides communication infrastructure based localization methods that map IP-addresses, GSM-cell identifiers or wireless router MAC-addresses to geographic locations, the main contribution to this development comes from the proliferation of GPS-enabled mobile user devices.

The critical point is that plain location data has the potential to both identify a single user and disclose sensitive information about that user's activity at the same time. This makes the robust anonymization of position information a non-trivial task and has created a lively branch in privacy research over the last years.

A useful first step in protecting the user's location privacy is a threat analysis, usually implemented by means of a privacy metric. However, most approaches presented in the literature (among them classical location privacy metrics like k-anonymity) depend on perfect knowledge about the state of the world (such as the positions of other users) which is unsuitable for many practical scenarios. Therefore we focused on a user-centric risk assessment – a privacy meter – which assumes the user to be autarkic, meaning that she does not hold any knowledge of the global system state. In fact it aims at providing the user with an estimate of her current location privacy level on the basis of directly available data sources like map data (e. g. from the OpenStreetMap project) and knowledge about the own movement and position disclosure history (available from local records). Our approach takes an attacker's perspective as a starting point and comprises several distinct indicators that estimate different aspects of the user's current threat level. Among these indicators are the street coverage of the region the user is located in, the street distribution in that region, the area reachable to the user in a certain time span, a map-based estimate of the people density in the region, and the number and reachability of points of interest in the vicinity of the user. The information a possible attacker can infer from past position disclosures of the user, such as the reachable area and maximum time spent at points of interest between two published positions, is also analyzed.

We tested this approach by simulation on real-world data, which yielded promising results about the benefits of the developed features. For some example user movements the indicators clearly distinguished between situations, where the user is more exposed due to a low diversity of possible activities (e. g. rural areas), and those where it is easy to hide in a crowd (e. g. a downtown area of a city). Especially calculating the implications of the spatio-temporal constraints given by two successive position updates resulted in interesting findings: In some of the example scenarios the actual route taken by the user could be reconstructed precisely, although the user's position was disclosed only every five minutes. Informing the user about this kind of consequences of a potential location publication might improve a privacy-aware usage of location-based services.

Furthermore we formalized these indicators and integrated them into a more comprehensive framework for a user-centric location privacy metric, taking different trust levels of communication peers (e. g. social contacts vs. location-based service providers) into account.

Open questions we are planing to tackle in the future comprise the use of location data in online social network services (e. g. geo-tagged status updates, posts and pictures). We expect the problem of inference attacks to be aggravated in this context where friends with a high background knowledge (e. g. about a constrained set of plausible locations) can take the role of attackers that may succeed with attacks even on only sparse location data.