Location Based Privacy In the world of Internet of Things and Big Data Analytics

By
Mohitkumar Rangholiya
School of IT, Deakin University

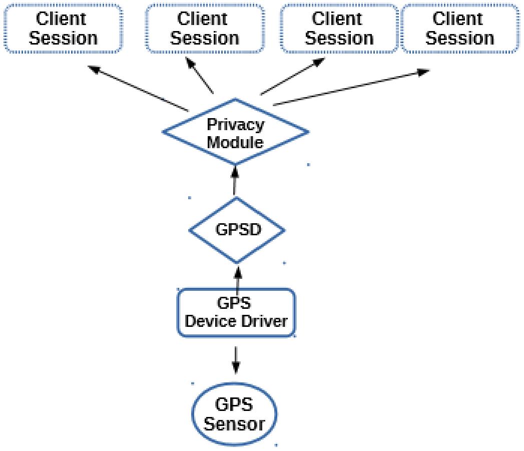
Location Based Privacy In the world of
Internet of Things and Big Data Analytics
68°F

Objective

To identify the need of protecting Data owner's location information and research the possible mechanisms that can be implemented in the direction of having granular control on data-owner's location data.

Methodology

In the era of smart devices ,smartphone owner's personal smart devices are continuously being tracked and regulated by many third party mobile applications and that is just because of no or less granular control of the data owners over their location information and hence the trust is constantly being violated. But mobile users sometimes want to share their location data with some granular over it. While third party application authorities can access data owner's location data which is not in the favor of data owners. Some researchers have tried to make some improvement in user's control by permissions(Felt et al. 2012), by some user defined constraints. Moreover permission manager in android and iPhone only allow users to either enable or disable the control. But complete fine granularity is still missing. The modified version of android operating system which is CynagonMod has also a concept named XPrivacy which allows data owners to configure random or static location for the specific application. It has lot of other location privacy modules that controls the release of location data from the device('XPrivacy') . All in all, it provides control over location data at application layer.



In this section ,the implementation of the privacy module to GPSD is presented. The below figure demonstrates the overall view of the queries and responses. It shows the privatization takes place before its release to the applications. This privacy module assures that the location data is released from the device only on user's control and permissions. Here some logical methodologies are mentioned to ensure the strong location privacy.

1. A Privacy Module which can be integrated GPSD software and runs on every GPS enabled device.
2. A granular privacy manager to control the location information release.
3. A performant privacy module with minimal overhead.

Here it's important to learn the architecture of GPSD. Then we can move further with the privatization algorithm and then we will go through the integration of privacy module with GPSD and then at the end we will evaluate the system.

Architecture : GPSD event loop

The architecture of GPSD components is presented here which shows the event loop. This event loop consists of accepting new client connections ,accepting client subscriptions and GPS reporting for clients. Every time when third party clients try to connect ,it has pass through the application identifier which is mapped to the system privacy configuration. System privacy configuration module contains different privatization radius in meters and random response coin flips. Then connection request is allowed to pass through the recommended set of privacy parameters but here it is also checked in parallel with the user defined privacy settings and it is not allowed to exceed the user's privacy threshold(Joy, Le & Gerla 2016).

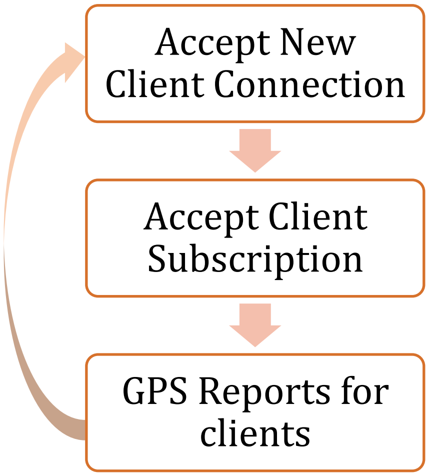


Figure 1GPSD Event Loop with placing Privatization when reporting to the clients

Privatization : LocationSafe

LocationSafe is a privacy module used over here and it supports major two types of privatization methods for user's location privacy. These privatization methods includes radius privacy and grid privacy using differential privacy.

Radius privacy method works on the user defined radius to define the random location within that radius. Here, the control of radius is in the hands of data owners so it can be clearly determined that the strength of the location privacy is ultimately defined by user itself. The larger the radius ,the stronger privacy can be achieved.

The another method called location grid privacy consider the location space as a location grid and the size of grid is achieved from the user defined specifications. Here, this location grid is built using the data owner's current location as one of the grid space and then using randomized response mechanisms ,one or more grid locations are sent with the response of location access request. This randomized response mechanism is depicted in the figure given below.

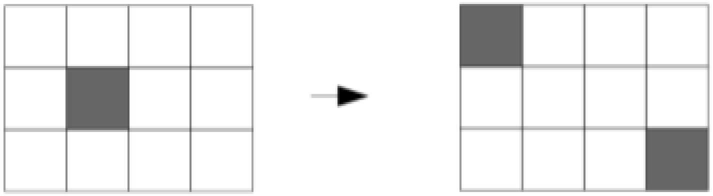


Figure 2Randomized response mechanisms using location grid

Here, the most useful mechanism of randomized response was firstly introduced by the social scientist to study sensitive attributes(such as use of drugs) of the population. In our method, the use of randomized location permits users to randomize their actual location and send these random location in the response of analyst's queries. After all, it satisfies the need of differential location privacy and ensures the optimal sample complexity for the differential privacy method. This way, it also guarantees the strong location privacy to the data owners.

Mathematical Logic : Randomized Response

owner reacts honestly; something else, the data owner flips the second coin and reports the consequence of this second coin flip. Assume there are N users taking an interest in the study. Let's Y consider as the aggregate total of "yes" randomized answers. The YA which is the estimated population considering sensitive attributes can be calculated.

The intention behind the idea of randomized response is that it gives "plausible deniability", i.e., any honest answer can create a reaction either "yes" or "no", and users hold powerful deniability for any answers they react. In the event that the first coin all the time comes up heads, there is high utility yet no security. On the other hand, if the first coin is always tails, there is low utility however high privacy. It has been demonstrated that controlling the two coin flips carefully, one can strike a harmony amongst privacy and utility.

Multiple Locations in Grid Privacy Mechanism

While considering randomized response mechanisms discussed so far is a powerful location privacy mechanisms for a solitary location-point, here the question raised is that how can it deals with the multiple locations, i.e., a matrix/grid representation? A deep study of "polychotomous" methods have been concentrated on and reviewed in the research paper(Fox & Tracy 1986) utilizing numerous randomizing components or greatest probability estimators(Tamhane 1981). However, it turns out that simply repeating an application of (Fox & Tracy 1986) for each grid location turns out to be an "optimal" (Tamhane 1981) mechanisms.

Example

After having above discussion & study, it can be expected that LocationSafe repeats the randomized response calculations for each & every grid locations. For better understanding of how LocationSafe does this can be achieved by considering an example. Suppose, a traffic analyst tries to analyze the flow of the traffic at a particular location. To carry out this, firstly he will generate a query asking the location information to all the nearby data-owners for their current location. Then after each data-owners will respond to this query by executing randomized response mechanism for their current location and revert with a Boolean bit vector(Joy, Le & Gerla 2016). Now, after having response from all the data owners, traffic analyst will perform aggregation of all received bit-vectors which will determine the traffic at that location considering the number of vehicles.

Limitations

- The randomized response privatization mechanism with grid privacy ensures strong privacy only by theoretical calculations not practically.
- The discussed mechanism needs user's input to define the level of privacy. So user should have little technical knowledge to define the level of privacy.
- LocationSafe is targeted to be implemented in core levels of operating system so implementation with android and iPhone can be difficult.
- It only focuses on the privacy of locations not the whole data privacy of IOT.

Ethical Issues

- Losses and Violations of Location Privacy caused by Location based technical systems. It means that it's not the issue when location sensors or GPS detect any individuals but it become ethical when it violates that location data. It is ethically significant to understand the difference between the losses and violations of privacy(Wang & Loui 2009).

Future Research Directions

- LocationSafe can be integrated with Android and iOS mobile operating systems. Using that one can evaluate the impact and design the user-interface for needed user-inputs.
- Strong security algorithms can be combined with privacy module in LocationSafe to improve the level of privacy of location data.

References

Felt, AP, Ha, E, Egelman, S, Haney, A, Chin, E & Wagner, D 2012, 'Android permissions: User attention, comprehension, and behavior', in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, p. 3.

Fox, JA & Tracy, PE 1986, 'Randomized response: a method for sensitive surveys'.

Joy, J, Le, M & Gerla, M 2016, 'LocationSafe: Granular Location Privacy for IoT Devices', *arXiv preprint arXiv:1606.09605*.

Tamhane, AC 1981, 'Randomized response techniques for multiple sensitive attributes', *Journal of the American Statistical Association*, vol. 76, no. 376, pp. 916-23.

Wang, JL & Loui, MC 2009, 'Privacy and ethical issues in location-based tracking systems', in *2009 IEEE International Symposium on Technology and Society*, pp. 1-4.

'XPrivacy'.