Advanced Topics on Privacy Enhancing Technologies
CS-523
Location Privacy Exercises

1 Who can track me?

Note: This exercise has many possible solutions, we provide in the solutions examples of what can be correct. You can come up with your own solutions and discuss with the TAs.

A big mall has built an Android app which recommends shops to users based on their interests. Initially, the application used GPS to locate users inside the mall. Unfortunately, GPS signals are weak inside buildings and their accuracy is not adequate for locating users inside a mall. To solve this problem, the app developers decided to use nearby Bluetooth and wireless signals to localize users. In this approach, location service providers (LSPs), such as Google, gather the location of wireless access points and Bluetooth beacons into a dataset; Whenever smartphone users want to know their location, they scan for nearby devices and send this data to an LSP to pin-point their location.

After reaching high accuracy on the raw location, the application monitors the location in the background every 30 seconds and sends it to the server to detect which shop the user is visiting. The application always keeps Bluetooth and wireless sensors on for better accuracy. After getting privacy complaints from users, the mall decides to sample a fresh noise for each reading and perturb the location in the app before sending the raw location to the server.

Evaluate the privacy of users and the impact of the location perturbation mechanism against the following adversarial models:

1. The mall which controls the application.

   Solution:
   The mall can track users' location with the application. The privacy gain of location perturbation depends on how much noise has been added to the data. If the noise is low enough that the application can detect the visited shop then the defense is not effective. However, if the noise is high enough to prevent detecting visited shops then the utility of the application will be drastically reduced. It's important to note that even under noise if a

user stays in a shop for a long time and the mall receives many perturbed locations from inside the shop then it can combine noisy locations to infer the visited shop.

2. The location service provider which localizes the user.

   Solution:
   The LSP learns the location history of the user. The client applies the perturbation after querying the LSP and getting the raw location. Hence, this defense has no impact on privacy towards the LSP as an adversary.

3. A shop which wants to detect frequent visitors.

   Solution:
   Many Bluetooth devices have unique identifiers. The shop can set up a Bluetooth receiver and log all visitors. The shop can track the duration and frequency of customers visits. The shop does not rely on location data from the app that is why perturbation has no impact on privacy.

4. A customer who jogs in the mall.

   Solution:
   Similar to the shop scenario, a customer can use a Bluetooth device to log nearby users. A human can see and recognize nearby people, so a single customer tracking does not have a high privacy impact. However, combining customers data together or with a central service can increase the risk/damage of this attack.

How can you improve the privacy for each adversary mentioned above?
Solution:

1. The mall: Instead of sending raw locations to the server and learning users' preferences inside the server, the application can push the POI detection and recommendation generation to the client-side and prevent sending location data to the server.

2. The location service provider: BLE/Wireless datasets are huge and they change over time, so having the full dataset in the client is not practical. However, if users have an inaccurate location (from GPS signal or knowing that the user is inside the mall), they can retrieve nearby beacons and use them to improve the accuracy of the localization. This approach reveals an inaccurate location to the LSP. It's possible to use generalization or perturbation to reduce the privacy risk in sharing this location.

3. A shop/customer: There is no need to use and advertise a unique wireless/BLE id for each smartphone. The smartphone can randomize this id periodically. Android 8 enabled MAC randomization capabilities and Android 10 enabled this option in the default setting. Apple's ios has similar defenses too.

2 k-anonymity cloaking

Clark is a resident of Smallville, a farming village with a low population density. He has developed SuperApp, a service to recommend sporting facilities to users based on their location. Clark decides to implement a k-anonymity cloaking scheme for SuperApp. The service works as follows: Users have SuperApp installed on their phones. SuperApp sends their location and query to a trusted third-party location anonymization service (LAS). The LAS computes a cloak. A cloak is an area based on the user's location that also contains $k-1$ other users. The LAS sends the cloak and the query to an untrusted location service provider (LSP). The LSP computes the results for the query and sends it back to the app. Clark is the developer, and can modify the code of SuperApp as well as provide the k value required by the LAS. He cannot modify anything on the LAS or the LSP. Assume that Clark is not malicious.

1. Clark needs to choose an appropriate value of k for his service. What would be a potential issue if Clark chooses a very high value of k?

   Solution:
   Since Smallville has a low population density, picking a high value of k could result in a cloak that spans a large area. This might reduce the utility of the app, by suggesting facilities that are not very close to the user.

2. Clark finally picks an appropriate k value for his service. He then tells his friend Lois about it. Clark and Lois decide to attract more customers by introducing SuperApp in Lois' city, Metropolis. Lois, who is in charge of releasing it in Metropolis, looks at the k-anonymization logic used by Clark. She concludes that since Clark has spent so much time fine-tuning the k value, it would provide sufficient privacy for the users of Metropolis. Is Lois correct in her conclusion? Justify.

   Solution:
   Metropolis has a larger population density than Smallville. Setting the same k value could result in a cloak of a smaller area. If that area spans a single building, for example, the LSP can infer that users are located in that particular building.

3. Clark decides that trusting the LAS is risky – if the LAS gets compromised, location data of the users could be revealed. He develops a workaround: the app creates dummy locations and sends them in addition to the real user location. This results in the cloaking area being calculated based on the user location and dummy locations (instead of other user locations in the system), and makes it harder for the LAS to determine the real location of a user. What are points that Clark has to keep in mind while implementing this change?

   Solution:
   Clark has to take the distribution of locations into account, so that users

get utility and privacy. For example, generation of locations that are spread over a large area could result in lower utility, whereas locations spaced very closely together might result in lower privacy. In addition to this, Clark has to develop dummy locations that are plausible and mimic actual user behavior, so that the LAS cannot distinguish real and dummy locations. For example, a location in the middle of a water body has a very low probability that it is an actual user location.

3 It's all a hoax

In a campaign to relax privacy regulations, a politician stated that "location privacy is just a big hoax". To prove his point, he publicly released the history of his location visits. As a firm supporter of strong privacy protections, you decide to prove him wrong and demonstrate in a short blog post how much sensitive information one can infer from an individual's location history. Describe shortly how you would go about inferring the following information from the politician's data public data:

1. Where does the politician live?

   Solution:
   The place where he most frequently spends between 1:00–5:00 AM is most likely his home. If he usually spends weekends and holidays in a rural place then it's possible that he has a summer house.

2. Where does he work?

   Solution:
   The place where he most frequently spends his 9-5 on weekdays is most likely his workplace.

3. Is he religious? If yes, what is his religion?

   Solution:
   Does he go to churches, mosques, Buddhist temples, etc? Frequency of his visits is a measure of his devotion.

4. Which companies had dealing with the politician?

   Solution:
   Companies that the politician has visited in the past year are more likely to have a business relationships with him. If companies' managers follow the politician's trend and release their location, then it's possible to detect off-company meetings, dining together, sharing common clubs, etc.

5. Does he have a healthy lifestyle?

   Solution:
   Check location history for jogging, biking, and outdoor activity. Compute his frequency of visiting gyms. Check how often he eats in restaurants.

Check visited restaurants for his food preference: fast food, vegan, steak house, ...

6. Is he healthy?

   Solution:
   Check visits to hospitals and clinics. If the accuracy is high enough, find the visited doctor and check his/her specialty. If the politician has a health condition it's likely that he see a doctor about it.