Initial Plan

Generating Fake Travel Source and Destination Location Pairs

Cardiff University

School of Computer Science and Informatics

CM3203 One Semester Individual Project: 40 Credits

Author: Andrei Cojocariu

Supervisor: Dr Padraig Corcoran

Project Description *3*
Project Aims and Objectives *4*
Work Plan *5*
Bibliography *7*

Project Description

Route planning services, such as Google Maps, are used by billions of people around the world (Reid, 2020) to obtain travel directions from one location to another. These services require users to submit source and destination location pairs. For example, if one wanted a route from their home to a local cinema, they would submit the location of their home and the location of the cinema. The act of submitting personal locations (e.g. the location of your home or workplace) to an independent service provider invokes a question of privacy. As expressed by (Corcoran, Mooney and Gagarin, 2020), "If one assumes the use of a secure communication protocol, such as SSH which exploits advances in encryption, the threat posed by an attacker eavesdropping on the communication can be considered minimal. Therefore the most significant threat is that posed by an attacker in the form of a service provider".

If a user is concerned about their privacy, they may decide to obfuscate or hide the true locations in question. This can be achieved by submitting to the service provider a set of fake source and destination location pairs along with the true location pair. When supplied with multiple location pairs, and the service provider cannot differentiate between true and fake pairs, they cannot determine which is the actual true location pair.

Therefore in the context of privacy the service provider cannot determine, when given a source destination pair, if the locations are anomalous (and therefore should be discarded). Even if fake data is kept and used by the service provider to build a profile of the user, the true user information will not be easily determinable.

The goal of the project is to develop a system for generating fake location pairs which cannot be distinguished from real location pairs corresponding to a given user. In order to create fake location pairs, we must first identify what we want to achieve with a 'fake' location pair. The real data locations that we want to obfuscate can be split into two categories: Repeated locations (Home, workplace, healthcare provider), and spontaneous locations (restaurants, nearest petrol station, nearest corner shop). Repeated locations are useful for the service provider in building a user profile, and spontaneous locations have very little value. Therefore, if the user is sending multiple real queries, the fake queries must consist of the same number of repeated and spontaneous locations as the respective real queries.

The solution proposed is a system that locally keeps track of user queries, building a profile consisting of the user's repeated locations and categorising them (e.g. home, workplace, frequent amenity). The system will then create a number of similar profiles, with the same categories, but with different locations that fit their respective categories (e.g. **Frequent Amenity: 'Starbucks'** is replaced with **Frequent Amenity: 'KFC'**). When the user sends a query, a query is sent simultaneously for each local profile. If a real location source or destination is identified as spontaneous, the fake profiles also use a spontaneous location in their respective queries. These fake profiles can then each be set in geographically distant locations to further protect the user's privacy.

Project Aims and Objectives

The objective of this project is to produce such a system in which the user inputs a source-destination location pair, and as the system is used over time, the fake source-destination location pairs are consistent in such a way that they are always indistinguishable from the real source-destination location pairs for that user.

The main aim of the project is for the system to allow input and output in a way that it can be used in conjunction with major route planning service providers, improving user privacy and reducing the accuracy of user-specific advertisements

Producing this system has three major milestones:

- System can create a single set of fake location pairs given a real location pair (complete by week 5)
- System produces consistent fake location pairs from a sequence of real location pairs (complete by week 6)
- System user interface allows for input of real location pairs, and full control of stored profiles (complete by week 8)

There will be two review meetings with my supervisor in this project: One during week 5 after the first milestone is complete, and one during week 9 when the system is complete and I have created a plan for the final report.

Work Plan

Key:

- Task

○ Milestone

- Review meeting with supervisor

<u>Week 1 - 31st January</u>

- Write initial plan
- Research appropriate query package/library for obtaining OpenStreetMap data

<u>Week 2 - 7th February</u>

- Research possible queries to be used for fake location pairs

○ Submit initial plan

<u>Week 3 - 14th February</u>

- Write queries to obtain real location pair data from OpenStreetMap
- Write query to obtain fake source location from a given real location pair
- Research possible methods to locally store and edit profile data

<u>Week 4 - 21st February</u>

- Write query to obtain fake destination location from a given real location pair
- Work on implementing user profile functionality
    A. Option to reset user profile
    B. Inputting a location pair revises the user profile

○ System can create fake location pairs from a given real location pair

<u>Week 5 - 28th February</u>

- Work on implementing fake profile functionality
    A. Option to delete fake profile
    B. Inputting a location pair revises all fake profiles

- Review meeting with supervisor

- Research appropriate input/output methods for user interaction
    A. GUI
    B. CLI

- System produces consistent fake location pairs from a sequence of real location pairs

<u>Week 6 - 7th March</u>

- Begin implementation of chosen interface method: GUI or CLI

<u>Week 7 - 14th March</u>

- Complete user interface functionality

- System user interface allows for input of real location pairs, and full control of stored profiles

<u>Week 8 - 21st March</u>

- Testing the system and fixing any bugs. General polish

<u>Week 9 - 28th March</u>

- Plan final report

- Review meeting with supervisor

<u>Easter Break: 2nd April - 24th April</u>

- Overflow time if behind planned target

<u>Week 10 - 25th April</u>

- Begin work on final report

<u>Week 11 - 2nd May</u>

- Continue work on final report

<u>Week 12 - 9th May</u>

- Finish final report

- Project submitted

Bibliography

Reid, E. (2020). A look back at 15 years of mapping the world. *The Keyword*. Available at: https://blog.google/products/maps/look-back-15-years-mapping-world/ [Accessed 2 Feb. 2022].

Corcoran, P., Mooney, P. and Gagarin, A. (2020). A distributed location obfuscation method for online route planning. *Computers & Security*, [online] 95, p.101850. Available at: https://www.sciencedirect.com/science/article/pii/S016740481831071X.