Master’s Thesis: Vehicle Traces Re-identification

Karim Emara, M.Sc
Chair of Applied Informatics and Cooperative Systems

Background

Future vehicle is connected and cooperative, whereby its intelligent sensors, real-time control and communication networks enable cars to communicate together or with infrastructure. Connected and cooperative cars (aka. Vehicular Ad hoc Network (VANET) or Car2X Communication) open the door for more intelligent applications regarding safety, traffic efficiency and infotainment. Most of these applications will depend on frequently sharing the current state of a vehicle such as a precise location, speed and heading. However, sharing this information should be carefully handled since it may threaten the driver’s privacy. Such movement information, when collected and analyzed, can expose sensitive facts about an individual, such as medical conditions, business connections or political affiliations.

A common privacy scheme is to use variant (in contrast to fixed) identifiers which is called pseudonyms to prevent an adversary from tracking and re-identifying the vehicle drivers. However, it was shown in [1] and [2] that tracking messages of variant pseudonyms is practically possible. Although vehicle tracking is crucial in privacy breaching, the adversary have to re-identify these anonymous traces to achieve a real privacy threat (i.e., correlate them to individuals). There are several re-identification techniques proposed such as [3, 4, 5, 6], but most of them assume that each vehicle uses a fixed pseudonym for all of its traces which facilitates the re-identification process.

Description

In this thesis, you will work on re-identification of vehicle traces of variant pseudonyms. In fact, this problem will be resolved if you can group traces traveled by each vehicle together. This can be accomplished by using a clustering technique on selected features of the traces (e.g., source/destination, frequency over week/month, followed routes, driving characteristics such as average and maximum speeds, etc.). Initially, you will work on a case where only one pseudonym is used by a trace. Then, you will work on the case where there are several pseudonyms used in each trace. In this case, you will use a tracker which can connect segments of different pseudonyms. This tracker is already implemented and tested and you do not need to understand its operation.

Methodology

The following aspects should be studied during the thesis:

- Investigate how vehicle traces can be grouped to their originating vehicle (Clustering).
- Evaluate and enhance the clustering technique when vehicle trace has variant pseudonyms.
• Develop a re-identification technique similar to ones used in references [3, 4, 5, 6].

• Obtain (or generate) realistic vehicle traces that reflect navigation among true drivers activities and point of interests (e.g. home, work, shopping mall, etc.) (many of real traces online. Search for: crawdad traces).

Prerequisites

• Practical skills in Data Mining and/or Machine Learning

• Good programming skills using Java, C++ or Matlab

• Self-motivated, Innovative and Independent

Application

If you are interested and really motivated, send your CV and transcript to me on emara@in.tum.de

References


