Preserving Traveler Privacy for Analyzing Repetitive Travel Patterns in Public Transportation Systems

Nadia Shafaeipour[1], Maarten van Steen[2,*], and Frank Ostermann[3,*]

[1]Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, the Netherlands
[2]Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), Digital Society Institute (DSI), the Netherlands
[3]Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, the Netherlands

Correspondence: Nadia Shafaeipour (z.shafaeipoursarmoor@utwente.nl)

Abstract. The widespread use of smart cards in public transportation can lead to the monitoring of travel behavior through unique identifying numbers, which raises concerns about privacy violations. To address this issue, this research proposes the use of Bloom filters to preserve the privacy of identifiers. We present some counting queries in this study, such as the number of travelers between an origin and destination, without being able to identify the travelers. The core of our research is determining what queries can be formulated and then dealt with without affecting accuracy substantially

Keywords. Privacy protection, travel patterns, public transportation

1 Introduction

Public transportation systems have become a crucial component of modern societies, providing an efficient means of mobility for inhabitants and contributing to the reduction of pollution (Boreiko and Teslyuk, 2016). To achieve maximum efficiency, transportation planners need to have a detailed understanding of passenger travel patterns across various modes of transport, including buses, trams, and metros. This knowledge is especially important given the wide range of travel patterns that exist, from workday commuting to weekend leisure trips, which can be affected by events and weather (Axhausen et al., 2002; Schlich and Axhausen, 2003). Automated fare collection (AFC) systems using smart cards have become increasingly popular, and many public transportations rely on smart cards for access and billing. The data collected from these cards is a rich information source. Typically, smart cards contain unique identification numbers that allow detailed monitoring of travel behavior. This level of detailed monitoring, however, represents a serious invasion of privacy that is not only in conflict with ethics but also with the law. Therefore, the analysis of data obtained from the use of such cards is generally subject to strict regulations under privacy laws (Asadpour and Dashti, 2011), such as the General Data Protection Regulation (GDPR) (Voss, 2016; Georgiadou et al., 2019).

As a result, it has been necessary to utilize privacy-preserving techniques while ensuring total anonymity for each individual while still being able to count repeat travelers. While there are various techniques available to preserve privacy in geospatial settings, such as obfuscation and k-anonymity, their effectiveness may be limited by certain constraints.

For this purpose, we propose a novel method based on (encrypted) Bloom filters (BFs), i.e. probabilistic data structures supporting set operations. This offers statistical counts of travelers as the only accessible information while protecting the privacy of travelers during both storage and processing.

The objective of this study is to explore the use of Bloom filters to achieve the following goals: accurately count the number of travelers moving between an origin and destination, while preserving their anonymity, and to identify commuting patterns between multiple locations.

2 Method

A Bloom filter (BF) (Bloom, 1970) is a space-efficient probabilistic data structure for storing sets of elements and testing set membership. We use Bloom Filters as representations of sets of detections Bloom filters are binary vectors in which the array consists of m bits initially set to zero, along with k different hash functions, each hash function returning a single position in the vector. When an element is added to a Bloom filter, it is hashed by each function, and the resultant position is set to 1. The size of the set represented by a Bloom filter can be estimated when knowing only *k*, *m*, and the number of 1s in the set. Detections of card identifiers are stored in a Bloom filter, after which we perform a bitwise homomorphic encryption. This kind of encryption allows us to perform operations *without having to decrypt* a Bloom filter. Queries on traveling patterns are formulated as combinations of sets. Answers are given in terms of (the size of) a set. By first shuffling the entries of an encrypted Bloom filter before handing it out as an answer to a query, the only sensible operation left is to calculate the size of the returned set. That size corresponds to the size of the original set, as represented by a Bloom filter.

To illustrate, suppose we wish to know how many travelers moved between location A and location B. Keeping matters simple, we detect card identifiers at A, and subsequently store these detections in an encrypted Bloom filter, say *bfA*. Likewise, we collect detections at B, which are stored in an encrypted Bloom filter *bfB*. By computing the intersection of *bfA* and *bfB*, we obtain the set of travelers between A and B, again represented as an encrypted Bloom filter. By shuffling the entries before handing out the result to a party that can decrypt the result, we hand out a Bloom filter representing a set that bears no relationship with the original intersection, yet has the same size.

Much more intricate queries can be addressed this way, yet we know that, notably, computing intersections affect the accuracy of estimating set sizes. Our study focuses on investigating which queries we can sensibly pose that will not lead to an unacceptable drop in accuracy.

We consider the following patterns:

- Travel patterns on working days are stable.

- Travelers prefer to follow the same trajectory during a specific time on all working days.

In light of the above travel patterns, we developed some queries that we may be able to address with high accuracy:

- What is the number of travelers who traveled from a specific origin-destination like A to B during a particular time (9:00 to 10:00) on each working day of the week? Does the number of travelers remain relatively constant throughout the week?

We were able to address such above counting queries. We applied the technique to the Beijing subway data. As a result, we counted passengers who moved between an origin and destination with a high degree of accuracy.

- What is the total number of travelers who traveled from origin A to destination B during a specific time frame (9:00 to 10:00), on each working day (Monday to Friday) of the week? (The number of travelers appearing every working day during s time frame).

Currently, we are working on addressing such a particularly complex query that involves computing intersections between multiple sets(sets of five working days).

References

Asadpour, M. and Dashti, M. T.: A privacy-friendly RFID protocol using reusable anonymous tickets, in: 2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications, pp. 206–213, IEEE, 2011.

Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfüser, G., and Haupt, T.: Observing the rhythms of daily life: A six-week travel diary, Transportation, 29, 95–124, 2002.

Bloom, B. H.: Space/time trade-offs in hash coding with allowable errors, Communications of the ACM, 13, 422–426, 1970.

Boreiko, O. and Teslyuk, V.: Structural model of passenger counting and public transport tracking system of smart city, in: 2016 XII International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH), pp. 124–126, IEEE, 2016.

Georgiadou, Y., de By, R. A., and Kounadi, O.: Location Privacy in the Wake of the GDPR, ISPRS international journal of geo-information, 8, 157, 2019.

Schlich, R. and Axhausen, K. W.: Habitual travel behaviour: evidence from a six-week travel diary, Transportation, 30, 13–36, 2003.

Voss, W. G.: European union data privacy law reform: General data protection regulation, privacy shield, and the right to delisting, The Business Lawyer, 72, 221–234, 2016.