International Journal of Information Management 34 (2014) 682–688

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Passengers information in public transport and privacy: Can anonymous tickets prevent tracking?

Gildas Avoine[a], Luca Calderoni[b,*], Jonathan Delvaux[a], Dario Maio[b], Paolo Palmieri[c]

[a] Information Security Group, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium
[b] Department of Computer Science and Engineering, Università di Bologna, 47521 Cesena, Italy
[c] Parallel and Distributed Systems Group, Delft University of Technology, 2628CD Delft, Netherlands

ARTICLE INFO

Article history:
Available online 19 July 2014

Keywords:
Privacy
Public transport
Sensitive data management
Privacy preserving technology

ABSTRACT

Modern public transportation companies often record large amounts of information. Privacy can be safeguarded by discarding nominal tickets, or introducing anonymization techniques. But is anonymity at all possible when everything is recorded? In this paper we discuss travel information management in the public transport scenario and we present a revealing case study (relative to the city of Cesena, Italy), showing that even anonymous 10-ride bus tickets may betray a user's privacy expectations. We also propose a number of recommendations for the design and management of public transport information systems, aimed at preserving the users' privacy, while retaining the useful analysis features enabled by the e-ticketing technology.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

In 2013, East Japan Railway (JR East), the largest rail company in the country, announced the intention to sell to Hitachi corporation a large dataset of its passengers' travel histories (Geuss, 2013). This information has been gathered by JR East through its proprietary e-ticketing system, *Suica*. The company plans to anonymize these data by replacing sensitive information, such as names and addresses of card owners, with anonymous ID's. But is this enough to protect users' identities, and therefore their privacy? Historically, public releases of anonymized personal information have often proved to be dangerous for the privacy of the people that information concerned. In 2006, America On Line (AOL) released anonymized data regarding the search queries of millions of users of its web search engine. Even if the IP addresses of the users were replaced by anonymous identifiers, researchers and even journalists had little trouble finding the real names of people corresponding to the anonymous ID's, as proved by the famous case of user #4417749, presented in a New York Times article (Barbaro & Zeller, 2006). In this paper we show how the disclosure of travel histories can be equally dangerous, as travel data contain a great deal of information about the user, even when her real identity is concealed by means of anonymization.

The use of electronic tickets, usually smart cards, in public transportation networks has a number of potential benefits, both for the users and the provider of the transportation service. Often, their introduction also coincides with a more general modernization of the transportation infrastructure. Modern networks usually integrate a positioning system (GPS) for monitoring the movements of buses and trams, backed by a constant Internet connection to a central control infrastructure. Enabling location-awareness allows, for instance, to display real time information and waiting times for each line on the provider's website or on information screens at bus stops and represent a value-added *city-to-citizen* service in the smart urban ecosystem (Calderoni, Maio, & Palmieri, 2012). Internet communication between vehicles and a central server can also be used to signal traffic congestion or unexpected issues efficiently, in both directions. These innovations help in making our cities smarter and greener, by improving the quality and reliability of the public transport service.

However, the technology enabling these features also generates an unprecedented amount of information regarding user movements. And after such information is generated, the tendency among public transportation companies is to record it, rather than discard it when it exhausted its original goal. In this paper we discuss the privacy implications of such a large amount of data, and we analyze the potential consequences of its disclosure. As electronic tickets are generally characterized by a unique ID, and all trips are recorded, the information stored in the information system of a public transport company is nothing less than a detailed log of each user's movements and therefore should be treated as

* Corresponding author.
E-mail address: luca.calderoni@unibo.it (L. Calderoni).

http://dx.doi.org/10.1016/j.ijinfomgt.2014.05.004
0268-4012/© 2014 Elsevier Ltd. All rights reserved.

sensitive information. However, in this paper we show that even when personal information of the users are not stored in the system – or are anonymized – the threat to privacy remains. In fact, combining the data in the transportation company database with other publicly available source of information can ultimately be enough to identify a specific user, even in the case of anonymous tickets.

1.1. Contribution

In this paper we focus on anonymous, disposable 10-ride electronic tickets for public transportation. Such tickets can be generally bought anonymously through resellers or automated machines and, while they are identified through a unique ID, they do not carry any information on the owner's identity. Thus, these tickets are perceived as the most privacy-friendly by the users while, at the same time, retaining some of the advantages of personal travel passes, such as a lower cost per ride than single-ride tickets, and the ability to be used multiple times. The choice of anonymous tickets allows us to evaluate the potential effects of disclosure of travel histories to third parties, even when limited to a small number of rides and anonymized by removing personal information.

In this paper we present the case study of a real, city-wide public transport network in Italy. By analyzing and decoding the tickets issued by the company, we infer the information collected during their use. We use this knowledge to show that even anonymized and numerically limited travel histories are indeed enough to profile users with a great depth of detail. We also show that careful elaboration of these data, and comparison with other publicly available sources of information ultimately allows to find matching patterns and to statistically identify the user as belonging to a small, well-defined group. Empirical evidence produced by analyzing this case study proves that simple anonymization of the travel histories of public transportation users is not sufficient to protect their privacy, and therefore suggests caution in the disclosure or trade of such data without the informed consent of the users themselves. In order to address this issue, we propose a set of recommendations for the design and management of the information systems of transportation companies. Our solutions are both privacy-preserving and cost-effective, as they reduce the overhead in communication and storage of travel data to the information system, while avoiding costly renovations of current infrastructures.

In this paper we focus on a specific case study (the Italian city of Cesena and its public transport system) and we analyze the potential information disclosure for a specific set of users (university students). However, the problem we bring to light is indeed common to other cities and countries. If in this work we use public information on students' classes and housing, the same result could be achieved, for instance, using phone directories. Overall, the aim of this paper is not to prove a flaw in the design of a specific e-ticketing systems, but rather to show how the disclosure or sale of location-aware information, such as travel histories, even when anonymous (or anonymized) could become dangerous to the privacy of the concerned individuals: in fact, when data are combined with other sources of information, the presumed anonymity disappears.

1.2. Related works

As discussed by Diaz and Gürses (2012), it is often difficult for individuals to know how their personal data are used by companies that hold them. While Diaz and Gürses mostly focus on sensitive data as defined by European Union regulations, their reasoning is equally valid when applied to companies collecting location data such as travel histories, as in the case of public transportation companies. In fact, users are often unaware of the risks of malicious surveillance, profiling or manipulation they are exposed to (Avoine, Martin, & Szikora, 2010). The security of this information is therefore best assured adopting the *Privacy by Design* paradigm, i.e. providing data anonymity by designing the appropriate protocols and procedures as hard-coded in the system itself (Diaz & Gürses, 2012). The privacy-friendliness of the infrastructure, if correctly implemented, does not necessarily hinder the business model (Liu, Bonazzi, Fritscher, & Pigneur, 2011). In the case of public transportation, electronic tickets raise privacy concerns for their ability to track users. Recent studies on the subject discuss this issue from the point of view of security against external attackers (Asadpour & Dashti, 2011; Avoine et al., 2010; Heydt-Benjamin, Chae, Defend, & Fu, 2006; Sadeghi, Visconti, & Wachsmann, 2008). A typical attacker is therefore some unauthorized person trying to monitor the movements of a victim, for instance by accessing the records of those movements stored on the ticket itself (usually a smartcard). For this reason, the studies usually conclude that no sensible information should be kept within the smartcard for longer than it is actually required for the correct functioning of the system. This is the case of Avoine et al. (2010), where the authors discovered, through an analysis of the Mobib smartcard (the public transport pass used in the city of Brussels, Belgium) the presence of unneeded information that could expose users to privacy threats. In this paper, instead, we are interested in privacy with respect to the company providing the transportation service. In a typical scenario of an RFID-based ticketing system, the smart card ticket is read on the vehicle in order to learn its unique identifier, which is then sent from the reader to the central server encrypted (Asadpour & Dashti, 2011), usually by applying a collision resistant hash function to the identifier. Unfortunately, this allows different stamps to be associated with the same user and therefore permits tracking (Sadeghi et al., 2008). In Kerschbaum, Lim, and Gudymenko (2013), the authors focus on electronic cash payments and bill processing in the e-ticketing scenario and discuss how to achieve a privacy preserving billing system based on asymmetric key encryption while in Peng and Bao (2010) a simple billing mechanism designed to avoid privacy leaks is proposed. Basically, it enables the public transport company not to collect the starting place and the ending one in order to compute the journey cost. Security and privacy issues related to *Near Field Communication*, a very common technology used for mobile-payments on public transports, are discussed by Salonen (2011). In Tseytin, Hofmann, Lyons, and O'Mahony (2006), the use of anonymous databases for collecting user movements is discussed. The authors show, from a theoretical point of view, that anonymity alone is not enough to protect users' privacy. In this paper we provide a real-world case study confirming their intuition, and we propose a set of plug-in privacy enhancements for existing information systems.

1.3. Laws pertaining public information

In this paper we show ways of de-anonymizing travel histories by comparing them to other sources of information. In order to show the viability of this approach, we use for this purpose only publicly available information. The case study we present focuses on Cesena's university students: we use therefore information from the local university dormitories and housing directories. In the following, we provide legal references showing how it is, in general, mandatory to maintain this personal information publicly accessible: this is due to transparency policies for applicants in merit rankings.

According to the Italian law D.P.R. 09.05.1994 n. 487 (published on the official gazette n. 185 on August 9, 1994), each ranking list related to a public competition must be published and accessible to the public. This is the case for instance of subsidized housing for students or public housing for disadvantaged people. More generally,

Download English Version:

https://daneshyari.com/en/article/1025674

Download Persian Version:

https://daneshyari.com/article/1025674

Daneshyari.com