Demo: Inspect what your location history reveals about you Raising user awareness on privacy threats associated with disclosing his location data

Antoine Boutet, Sébastien Gambs

Demo: Inspect what your location history reveals about you

Raising user awareness on privacy threats associated with disclosing his location data

Antoine Boutet
Univ Lyon, INSA Lyon, Inria, CITI
antoine.boutet@insa-lyon.fr

Sébastien Gambs
Université du Québec à Montréal
gambs.sebastien@uqam.ca

ABSTRACT

Location is one of the most extensively collected personal data on mobile by applications and third-party services. However, how the location of users is actually processed in practice by the actors of targeted advertising ecosystem remains unclear. Nonetheless, these providers have a strong incentive to create very detailed profile of users to better monetize the collected data. End users are usually not aware about the strength and wide range of inference that can be performed from their mobility traces. In this demonstration, users interact with a web-based application to inspect their location history and to discover the inferential power of this kind of data. Moreover to better understand the possible countermeasures, users can apply a sanitization to protect their data and visualize the impact on both the mobility traces and the associated inferred information. The objective of this demonstration is to raise the user awareness on the profiling capabilities and the privacy threats associated with disclosing his location data as well as how sanitization mechanisms can be efficient to mitigate these privacy risks. In addition, by collecting users feedbacks on the personal information revealed and the usage of a geosanitization mechanism, we hope that this demonstration will also be useful to constitute a new and valuable dataset on users perceptions on these questions.

CCS CONCEPTS

• **Security and privacy** → **Social aspects of security and privacy**; *Privacy protections*; *Usability in security and privacy*;

KEYWORDS

Location Privacy, Awareness, User Tracking, Mobile applications

ACM Reference Format:
Antoine Boutet and Sébastien Gambs. 2019. Demo: Inspect what your location history reveals about you: Raising user awareness on privacy threats associated with disclosing his location data. In *The 28th ACM International Conference on Information and Knowledge Management (CIKM'19), November 3–7, 2019, Beijing, China.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/de0207

CIKM '19, November 3–7, 2019, Beijing, China

ACM ISBN 978-1-4503-6976-3/19/11.
https://doi.org/10.1145/de0207

1 INTRODUCTION

As the profiling has become the norm on the Internet, the personal data of users is massively collected without the consent of the individuals concerned [5]. Due to the wide adoption of mobile devices, the location data of users is obviously part of the tracking and the most extensively collected data [2]. This tracking is usually performed through the usage of mobile applications exploiting the location of users. In these Location-Based Services (LBS for short), the position of the user is usually sent to a distant server, which process it to provide contextual and personalized answers or simply to store this information for profiling purpose (*e.g.*, sometimes the location is collected even if it is not necessary for the application). Although these LBS can provide useful information for users, once the data has been collected by a third party nothing prevents it from analyzing and possibly sharing the collected information for commercial purpose, which opens the door to many privacy threats. Examples of such leaks of personal information are regularly covered by news media and can include sensitive information such as the HIV status of the users (*e.g.*, Grindr [12]).

The providers of Internet services and mobile applications have a strong incentive to profile users based on the personal information collected and to monetize these profiles for targeting purposes. Indeed, the monetization of user profiles is the main source of funding for most of these providers. Despite the fact that web tracking has been a very well investigated field of study since almost 20 years, in the mobile context determining which information is collected and how it is processed and used remain a challenging issue [4]. As a result, this lack of transparency coupled with the emergence of controversial practices such as discrimination [9] raises serious concerns. A recent study shows that the most sensitive and valued category of personal information is location [17]. However, end users are usually not aware about this profiling as well as the type and the accuracy of the information that can be inferred from their location histories as well as the associated privacy and discrimination issues.

The privacy issues raised by location data have received quite a lot of attention in the last years. In particular, recent works have demonstrated that mobility is a very rich contextual information in the sense that it has a strong inferential potential in terms of information that can be predicted about the individuals whose movements are recorded. For instance,

researchers have shown that analyzing mobility traces can reveal personal data about individuals such as their points of interests (*e.g.*, home and place of work) [8], their race and gender [18], their social network [16] as well as to predict their mobility [15], to link accounts of the same user across different datasets [14] and to uniquely identify users from anonymous datasets or to conduct a de-anonymization attack [7]. Moreover, it is possible to analyze the semantics of these mobility traces to infer even more sensitive information such as their religion [6]. In addition, other studies have also demonstrated that the location is used for price discrimination [11].

The main objective of this demonstration is to raise the user awareness about the profiling capabilities related to the disclosure of their personal location data and the associated privacy and discrimination threats. More precisely, users are invited to analyze their location history collected and provided by Google for ensuring data portability (*i.e.*, one of the new right that has appear with the General Data Protection Regulation) and to inspect the information that can be inferred from the collected data. Then, we build and present to users a contextual profile combining location data with semantic information deduced from the mobility traces such as their points of interests, their home and work places, and the associted demographic information. Moreover, the rationale behind each information appearing in the contextual profile presented to users is also detailed to users so that can better understand how such inference was possible.

In addition of the inference attacks presented previously, an important literature has been devoted to developing protection mechanism for location data this last decade. However, none of the proposed Location Privacy-Preserving Mechanisms (LPPMs for short) have been adopted by mobile applications and the location data of users are still collected without any protection in real-life. In this demonstration, we will propose the users to apply an LPPM on their mobility traces implementing the privacy notion of Geo-indistinguishability [3]. Thus, users can both visualize the mobility traces as outputted by the LPPM and inspect the associated inferred information and contrast it with inferences performed on the raw mobility traces. Finally, users are also invited to provide feedbacks on the accuracy of the information shown as well to quantify their level of (un)comfort with the disclosure of these personal information. We ambition to use these feedbacks to create a new dataset containing the perception of users on location privacy with the hope that it can be used to design new protection mechanisms that meet their expectations.

In the following sections, we review the related work before describing our application and how users can interact with it to understand the profile that can be inferred about them, as well as to observe the impact of an LPPM on the mobility traces and the inferred information. Our demonstration is available at http://tiny.cc/c8591y where users can upload and inspect their own location history or analyse location data of a set of users. Attendees will be invited to use our demonstration which will display through a large screen available on our desk.

2 RELATED WORK

Only few tools have been proposed to inspect location data and study the expectation of users about their location privacy. For instance, Data Track [10] has investigated the perception of users about their right of data portability from service providers. This tool allows users to visualize locally on their machine the data export on their subject. However this tool is not devoted to location data and it does not performed any inferences. Google also provides a web interface in which users can explore their location history through a timeline [1]. More precisely, users are shown a list of places in which they made a significant stop, their different journeys as well as the associated transportation mode. Nonetheless, no attendance statistic on these places and no inference on personal information (*e.g.* home and place of work, gender) are provided.

FindYou [13] is another tool allowing users to inspect the potential of their location data. FindYou reports on a map the location associated to pieces of content with geolocalization metadata shared on social networks. Additionally, this tool predicts the home place and leverages on census data to infer demographics information. However, the demographics are only available for the USA and the location data available through the considered social media are considerably much more sparse than the location history of Google. Finally, our demonstration makes an important step forward by providing users the first tool to analyse the impact of LPPMs for preventing the profiling of their own location history.

3 DEMONSTRATION

In this section, we describe in more details our demonstration, first by presenting the incoming location data (Section 3.1) and then by reviewing the inferred contextual profile and how we build it (Section 3.2). Finally, we present how the location data is protected and the associated impact on both the mobility traces and the inferred profile (Section 3.3) as well as the conducted survey to capture user's perception on location privacy (Section 3.4).

3.1 Location Data

To better raise user awareness on the potential privacy threats associated with disclosing location data to third party services, we invite the users and the attendees of the demonstration to use their own location history. For enforcing the requirements pursuant to the GDPR (General Data Protection Regulation) in term of data portability and control over data, Google allows every user to export their data from different Google products such as email, calendar, photos and location history. The location data are collected from the Android-based smartphone as well as LBS. Consequently, we ask the attendees to export their location history from Google to feed our demonstration. Obviously, we explain users how to get their mobility history from Google and we

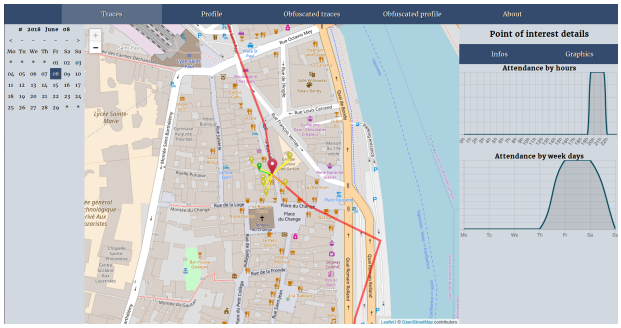


Figure 1: Visualization of a mobility trace and information about a detected point of interest.

ask them to agree our privacy policy to process and store their personal data for scientific purposes.

3.2 Contextual profile

Our demonstration implements some of the state-of-the-art inference attacks on location data. First, we identify the Points of Interests (POIs for short) of users by processing their mobility history. POIs are defined as spatially delimited places in which users spend chosen amount of time. For examples, POIs can be home or work places, but also a swimming pool, a school, a theater or a restaurant. In addition, they can also be even more sensitive such as a religious monument in which a user regularly goes, the headquarters of a political party she is involved in or a hospital she is treated in. POIs are usually extracted from mobility traces by using clustering algorithms like the ones presented in [19]. A POI is characterized by the diameter d of the observed location records and the duration t of the stop. Each granularity level (*i.e.*, value of d and t) reveals different information. For instance, a large and long stop (*i.e.*, a large d and a long t) can reveal that a user spends one day in a university campus but a shorter d and t can disclose the specific buildings in which the user spends time and possibly its specific department. By selecting a small t, it is also possible to detect the short stops of users. In the demonstration, we define a POI as an area of 150 meters of diameter where users spent at least 30 minutes.

We exploit open APIs to extract meaningful information about these POIs. More precisely, we exploit the OpenStreetMap API to find the address associated to coordinates of the POI, a picture corresponding to this address, and details about the associated place as shown Figure 1. A place can be of different categories (*e.g.*, a restaurant or a school) and includes a description. In addition, we report attendance statistics for each POI such as regularities and the temporal context (*i.e.*, the moment in the day or in the week).

Finally, by combining information about POIs, open API, regularities and the temporal context of the attendance, as well as information from census we are able to construct a contextual profile of the considered user (Figure 2). This profile encompasses 1) the home and a picture associated to the location (Figure 2b), 2) the working place with a description and a picture of the location, 3) the list of POIs including associated information, their regularities and statistics of the user's attendance, 4) a list of tags associated to the visited places, and 5) prediction of the gender (Figure 2a), the age and the salary. We also explain each piece of information in this contextual profile by showing the elements that have lead to this inference. For instance, a regular place where the user spends the night is predicted as a home place, and a regular place where the user spends the day is predicted as a working place. Additionally, by crossing information from reports of national statistics institute and both the home and the working places, we are able to predict the gender, the age, and the salary. Moreover, we ask users to provide a feedback on the accuracy of each information of their contextual profile.

3.3 Location privacy

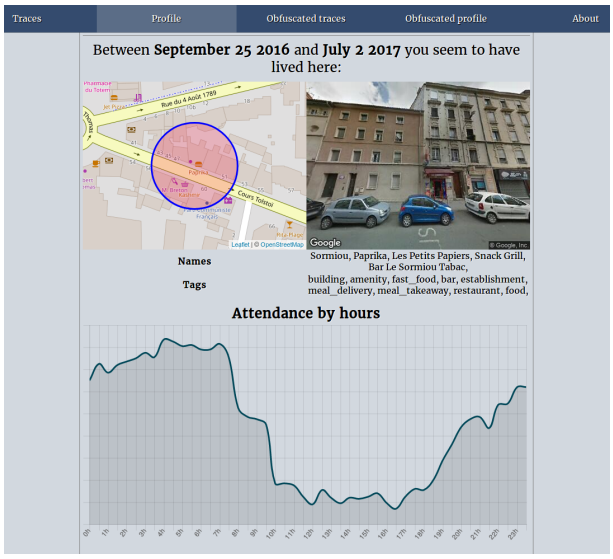
To protect users location data, many LPPMs following different privacy models have been proposed these last years. Due to a possible high drop of utility and its scalability issue, the k-anonymity model has now been replaced by differential privacy model as the main used privacy model. Nonetheless, users are generally not aware about LPPMs and the associated impact on their mobility traces. To raise awareness about these mechanisms and their potential in term of privacy improvement, in this demonstration we also let the users apply an LPPM on their mobility traces. More precisely, we apply an LPPM based on the privacy notion of Geo-Indistinguishability [3]. In its most basic form, this LPPM introduces a noise drawn from a planar Laplace distribution on each location. The amplitude of the injected noise is controlled by an epsilon parameter, in this demonstration $\epsilon = 0.05$. Users can visualize the resulting protected location data as well as analyze the associated inferred information. In addition, we highlight the information in the contextual profile that have been removed due to the protection mechanism (compared to using the raw data).

3.4 User survey

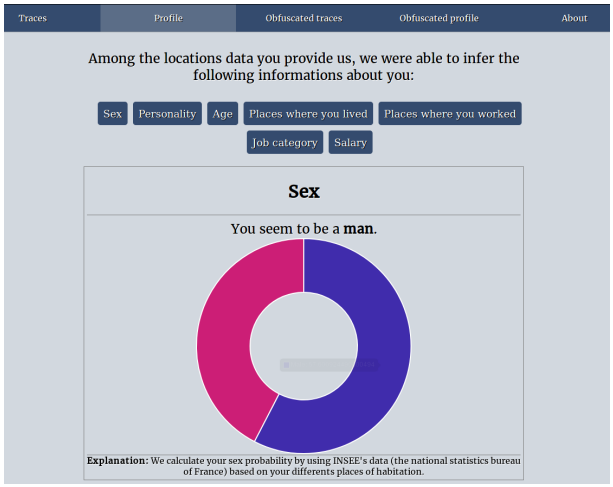
Lastly, we conduct a user survey by asking them how comfortable they are about this information disclosure and for each entry if they consider the information as sensible and that they would like that it remains private. We also ask the perception of users on the protection mechanism and its impact on the granularity of the predicted information.

4 CONCLUSION

Tracking has now become omnipresent on the Web, most often without the users knowing it and without their consent. With the widespread adoption of mobile phone including location capabilities, the location of user is obviously part of the tracking. The location is indeed a rich information that can reveal extensive information about the individual. In this demonstration, we propose to raise the awareness of users about the information that can be inferred from their location history. The attendees are invited to participate to the demonstration by using their own location history. By



(a) Working place



(b) Gender

Figure 2: From the mobility traces of the user, we build and show the inferred contextual profile.

analyzing these location data, we are able to provide to users an extensive contextual profile. Moreover, we show to users what would be their data by using a protection mechanism and the associated impact on the inferred information. One important aspect of this demonstration is also to collect a new dataset containing the perception of users about the information disclosure related to the exploitation of location data as well as the impact of LPPMs.

We aim to extend this demonstration by new inference attacks as well as new LPPMs. Specially, we plan to use additional API and bases of knowledge to feed our inference engine with more semantic information about the buildings to better infer the activities and interests of the users associated to their POIs. In addition, we implemented a novel inference attack predicting the big five personality traits of users from their characteristics derived from their mobility data using a supervised machine learning techniques (i.e., a multi-task regression algorithm). To have enough data to train our prediction model, we ask users to fill out a form to collect ground truth on their personality.

ACKNOWLEDGMENT

This work has been partially funded by the CHIST-ERA project UPRISE-IoT (User-centric Privacy and Security in the IoT).

REFERENCES

[1] [n. d.]. Google location history. https://www.google.fr/maps/timeline. ([n. d.]).
[2] Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. 2015. Your Location Has Been Shared 5,398 Times!: A Field Study on Mobile App Privacy Nudging. In *CHI*. 787–796.
[3] Miguel E. Andrés, Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. 2013. Geo-indistinguishability: differential privacy for location-based systems. In *CCS*. 901–914.
[4] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2015. Automated Experiments on Ad Privacy Settings. *PoPETs* 2015, 1 (2015), 92–112.
[5] Steven Englehardt and Arvind Narayanan. 2016. Online Tracking: A 1-million-site Measurement and Analysis. In *CCS*. 1388–1401.
[6] Lorenzo Franceschi-Bicchierai. 2015. Redditor cracks anonymous data trove to pinpoint Muslim cab drivers. http://mashable.com/2015/01/28/redditor-muslim-cab-drivers/. (Jan. 2015).
[7] S. Gambs, M. O. Killijian, and M. N. d. P. Cortez. 2013. De-anonymization Attack on Geolocated Data. In *TrustCom*. 789–797.
[8] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núòez del Prado Cortez. 2011. Show Me How You Move and I Will Tell You Who You Are. *Trans. Data Privacy* 4, 2 (Aug. 2011), 103–126.
[9] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. 2014. Measuring Price Discrimination and Steering on E-commerce Web Sites. In *IMC*. 305–318.
[10] Farzaneh Karegar, Tobias Pulls, and Simone Fischer-Hübner. [n. d.]. Visualizing Exports of Personal Data by Exercising the Right of Data Portability in the Data Track - Are People Ready for This?
[11] Jakub Mikians, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. 2012. Detecting Price and Search Discrimination on the Internet. In *HotNets*. 79–84.
[12] Brian Moylan. 2018. Grindr was a safe space for gay men. Its HIV status leak betrayed us. https://www.theguardian.com/commentisfree/2018/apr/04/grindr-gay-men-hiv-status-leak-app. (2018).
[13] Christopher Riederer, Daniel Echickson, Stephanie Huang, and Augustin Chaintreau. 2016. FindYou: A Personal Location Privacy Auditing Tool. In *WWW*. 243–246.
[14] Christopher Riederer, Yunsung Kim, Augustin Chaintreau, Nitish Korula, and Silvio Lattanzi. 2016. Linking Users Across Domains with Location Data: Theory and Validation. In *WWW*. 707–719.
[15] Adam Sadilek and John Krumm. 2012. Far out: Predicting Long-term Human Mobility. In *AAAI*. 814–820.
[16] Kumar Sharad and George Danezis. 2014. An Automated Social Graph De-anonymization Technique. In *WPES*. 47–58.
[17] Jacopo Staiano, Nuria Oliver, Bruno Lepri, Rodrigo de Oliveira, Michele Caraviello, and Nicu Sebe. 2014. Money Walks: A Human-centric Study on the Economics of Personal Mobile Data. In *UbiComp*. 583–594.
[18] Yuan Zhong, Nicholas Jing Yuan, Wen Zhong, Fuzheng Zhang, and Xing Xie. 2015. You Are Where You Go: Inferring Demographic Attributes from Location Check-ins. In *WSDM*. 295–304.
[19] Changqing Zhou, Dan Frankowski, Pamela Ludford, Shashi Shekhar, and Loren Terveen. 2004. Discovering Personal Gazetteers: An Interactive Clustering Approach. In *GIS*. 266–273.