Title:
Decentralized AI for Privacy Preserving Location-based Online Services

Keywords:
security, privacy, data protection, distributed/federated machine learning, online geo-located services

Advisor:
Sonia Ben Mokhtar

Co-advisor:
Sara Bouchenak

DETAILED DESCRIPTION

Geo-location data is increasingly used to improve the quality of services, leading to the surge of Location Based Services (LBS) such as navigators or nearest places recommendation applications. This generates very large amounts of mobility data that are currently used by companies and researchers. Indeed, the processing of mobility data can reveal many valuable information that may be used for a broad range of applications, e.g., traffic congestion management, urban development, etc. However, the processing of location data also comes with threats on the privacy of the recorded users. The most common threats are (i) re-identification attacks where the identity of an anonymous user is guessed based on previously recorded data, (ii) mobility prediction that anticipates users' next moves based on their habits, (iii) extraction of user's places of interest (home, workplace, etc.) and (iv) inference of social relationships (partners, coworkers, etc.).

To overcome these privacy issues, many efforts in the literature aim to develop protection mechanisms. The protection efforts are not only motivated by cautious companies and researchers but is more and more forced by national and international governments and organizations. The so-called Location Privacy Protection Mechanisms (LPPM) modify the location information of users to improve their privacy level. The principle behind each LPPM varies, for instance Geo-Indistinguishability (GEO-I) adds noises to the spatial information of the user data [1], PROMESSE modifies timestamps in order to smooth the user speed [2], and CloakDroid assigns the value of a location point to its closest location on a grid [3].

In this context, a problem that mobile app developers aiming at enforcing privacy-by-design have to solve is: **"how to objectively compare the privacy vs. utility tradeoff offered by different LPPMs and choose the right one ?"**

In practice, solutions that have been explored in the literature to select among a set of LPPMs generally rely on re-identification attacks [9]. Specifically, these solutions apply various LPPMs on a given trace and choose the LPPM (and its corresponding configuration) that better resists a given set of re-identification attacks. The role of these attacks is to link anonymous traces to past user data. However, to reach this objective, the proposed solutions assume a trusted proxy server as existing re-identification attacks are centralized:

they build user profiles using past unprotected mobility data and use them to estimate to whom a given protected trace belongs to.
To avoid centralizing raw data in a trusted proxy server, we aim in this PhD project to investigate the utilisation of the **Federated Learning (FL) paradigm**[10]. FL is a new machine learning technique which proposes to distributively train the models right where the data is created, i.e., on the user mobile devices.

The objective of this PhD project is to precisely address this question with the following objectives:

- Study the state of the art in terms of location privacy protection mechanisms, re-identification attacks and federated learning
- Design a novel risk assessment algorithm using a federated learning approach.
- Design a novel location privacy protection mechanism that relies on the above risk assessment mechanism
- Assess the privacy vs utility tradeoff of the proposed solution compared with competitors from the state of the art

References

[1] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Geo-indistinguishability: Differential Privacy for Location-based Systems. In CCS, pages 901–914, 2013.

[2] V. Primault, S. Ben Mokhtar, C. Lauradoux, and L. Brunie. Time distortion anonymization for the publication of mobility data with high utility. In TrustCom, pages 539–546, 2015.

[3] K. Micinski, P. Phelps, and J. S Foster. An Empirical Study of Location Truncation on Android. Most'13, 2013.

[10] K. A. Bonawitz Hubert Eichner Wolfgang Grieskamp Dzmitry Huba Alex Ingerman Vladimir Ivanov Chloé M Kiddon Jakub Konečný Stefano Mazzocchi Brendan McMahan Timon Van Overveldt David Petrou Daniel Ramage Jason Roselander. ACM SysML 2019

Selected publications of the advisors related to the topic

[4] M. Maouche, S. Ben Mokhtar, S. Bouchenak. HMC: Robust Privacy Protection of Mobility Data Against Multiple Re-Identification Attacks. ACM Journal on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(3), September 2018.

[5] S. Cerf, S. Bouchenak, B. Robu, N. Marchand, V. Primault, S. Ben Mokhtar, A. Boutet, L. Y. Chen. Automatic Privacy and Utility Preservation of Mobility Data: A Nonlinear Model-Based Approach. IEEE Transactions on Dependable and Secure Computing, 2021.

[6] R. Talbi, S. Bouchenak, L. Y. Chen. Towards Dynamic End-to-End Privacy Preserving Data Classification. IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2018), Fast Abstract, Luxembourg, June 25-28, 2018.

[7] R. Pires, D. Goltzsche, S. Ben Mokhtar, S. Bouchenak, A. Boutet, P. Felber, R. Kapitza, M. Pasin, V. Schiavoni. CYCLOSA: Decentralizing Private Web Search Through SGX-Based Browser Extensions. The 38th IEEE International Conference on Distributed Computing Systems (ICDCS 2018), Vienna, Austria, July 2-5, 2018.

[8] S. Cerf, V. Primault, A. Boutet, S. Ben Mokhtar, R. Birke, S. Bouchenak, L.Y. Chen, N. Marchand, B. Robu. PULP: Achieving Privacy and Utility Trade-Off in User Mobility Data. SRDS 2017. The 36th IEEE Symposium on Reliable Distributed Systems (SRDS 2017), Hong Kong, September 26-29, 2017.

[9] Besma Khalfoun, Mohamed Maouche, Sonia Ben Mokhtar, Sara Bouchenak:MooD: MObility Data Privacy as Orphan Disease: Experimentation and Deployment Paper. Middleware 2019: 136-148