Post-doctoral position

Protecting privacy in mobile apps using predictive control and machine learning techniques

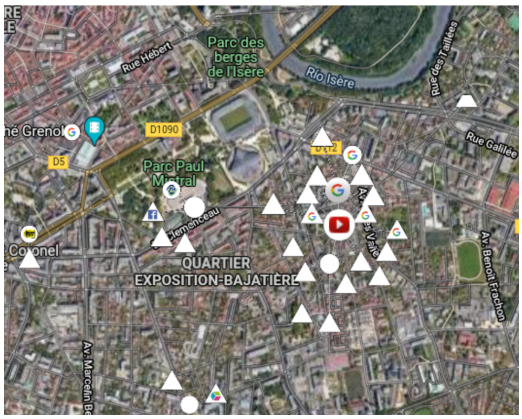
Supervisors:

- Mirko FIACCHINI, *CNRS, Gipsa-lab*
  mirko.fiacchini@gipsa-lab.fr
  Tel. +33 (0)4 76 82 62 25
- Sophie CERF, *INRIA center of the University of Lille*
  sophie.cerf@inria.fr
  Tel. +33 (0)3 59 57 87 62
- Bogdan ROBU, *Université Grenoble Alpes, Gipsa-lab*
  bogdan.robu@univ-grenoble-alpes.fr
  Tel. +33 (0)4 76 82 63 26

Institutions: This Post-Doctoral research is held in collaboration between the MIAI (Multidisciplinary Institute in Artificial Intelligence) Research Institute [1] and Université Grenoble Alpes, France

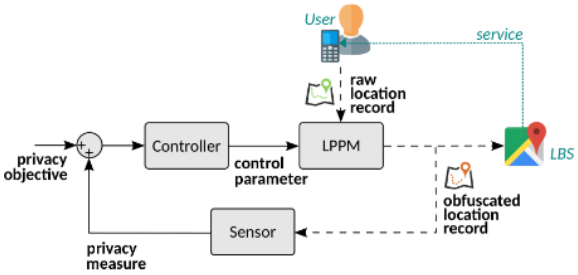
General context of the project: Privacy / Utility tradeoff in mobile tracking applications

Predominant in nowadays society, mobile apps are rising as application systems for control theory. Indeed, an app can be seen as a plant processing input signals and generating outputs. Ensuring that the app complies with a desired behavior, with guarantees, is a major safety challenge with impact on a very wide scope, as highlighted by the European Commission Strategic Plan [1] or the Location Privacy Protection Act [2] in the US.



Regulation of IT systems has emerged in the 2000's with the concept of Autonomic Computing [8], aiming at software self-adaptation. State-of-the-art works investigate the use of control theory for modeling and decision-making in computing systems, opening an entire promising field of research.

Scope of the research: Privacy in mobile apps



Location privacy protection considers mobile devices users whose mobility information is shared with untrusted parties. Applications and services tend to require location data to personalize users experiences. Examples of location-based services are navigation applications, recommendation systems, weather forecasting or fitness tracking apps [a].

[a] https://play.google.com/store/apps/category/TRAVEL_AND_LOCAL

Mobile apps provide those personalized and convenient services at the cost of personal data disclosure (one gains in service utility at the cost of sharing personal data). Service providers or third party attackers can take advantage of these data to derive private information about users.

Location Privacy Protection Mechanisms emerged as solutions to protect users privacy [12]. Such algorithms modify the location data to improve privacy, e.g. by adding noise [4], [6], reducing data precision [7], or merging close users locations [3].

[1] https://miai.univ-grenoble-alpes.fr/

In this research we focus on an online solution to improve performance, that is intelligently adding noise to the location data. The aim of the obfuscation methods is to hide the relevant information on the user position while preserving utility services quality by artificially perturbing the transmitted location.

Data is an important part of this project as the developed techniques will need to be validated on real location records. The team has access to four mobility datasets collected in a real environment: the Cabspotting[2], the PRIVAMOV[3], the GEOLIFE [14], and the Mobile Data Challenge datasets [9], amounting to a total of 770 users and more than 14 000 000 records.

Objectives of the post-doc: Optimization techniques, as well as optimal and predictive control methods [5, 13], can be used to infer and predict the optimal obfuscated location to be transmitted with the aim of maximizing the competing objectives of utility and privacy preservation. Machine learning tools [10, 11] can be used for the online implementation of low computational approximations of the optimal obfuscation strategies.

The obfuscation problem consists, in fact, in the online determination of the value of the position to be transmitted that maximizes the privacy among those that guarantee a minimal utility measure. Analogously, an optimization problem can be posed aiming at minimizing the utility loss among the possible transmissible positions ensuring a certain minimal privacy level.

Moreover, based on the available dataset, it is possible to compute offline the optimal obfuscated positions that minimize a cost concerning also the future values of utility loss and privacy performance. This would permit to take into account also the effect of the current choice in the future values of utility and privacy, introducing an additional predictive action to the obfuscation method.

Once the optimal predictive values are computed offline, the optimal obfuscation function could be estimated by using learning techniques with the aim of obtaining low complexity approximations of the optimal solution to be applied online.

Requirements: The applicant must comply to the following requirements:

- PhD in control or any related field
- Experience with nonlinear programming and optimization tools
- Skills with Python (Pandas, Sklearn) and Matlab.

Knowledge on IT, mobile systems and privacy will be appreciated but are not required.

General Information: Town/country : Grenoble, Gipsa-lab research laboratory, France
Starting date : as soon as possible, early 2022
Duration of contract : 1 year
Gross monthly salary (before taxes) : 2 656 €

References

[1] Orientations towards the first strategic plan for horizon europe. https://ec.europa.eu/info/files/orientations-towards-first-strategic-plan-horizon-europe_en. Accessed: 2021-10-06.

[2] S.1223 - location privacy protection act of 2012. https://www.congress.gov/bill/112th-congress/senate-bill/1223. Accessed: 2021-10-12.

[3] Osman Abul, Francesco Bonchi, and Mirco Nanni. Never walk alone: Uncertainty for anonymity in moving objects databases. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 376–385. IEEE, 2008.

[4] Miguel E. Andrés, Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Geo-indistinguishability: Differential Privacy for Location-based Systems. In *CCS*, pages 901–914. ACM, 2013.

[2] http://crawdad.org/epfl/mobility/20090224/cab/index.html
[3] https://projet.liris.cnrs.fr/privamov/project/dataset

[5] John T Betts. *Practical methods for optimal control and estimation using nonlinear programming*. SIAM, 2010.

[6] Sophie Cerf, Sara Bouchenak, Bogdan Robu, Nicolas Marchand, Vincent Primault, Sonia Ben Mokhtar, Antoine Boutet, and Lydia Y. Chen. Automatic privacy and utility preservation for mobility data: A nonlinear model-based approach. *IEEE Transactions on Dependable and Secure Computing*, 18(1):269–282, 2021.

[7] Chi-Yin Chow, Mohamed F Mokbel, and Xuan Liu. A peer-to-peer spatial cloaking algorithm for anonymous location-based service. In *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, pages 171–178, 2006.

[8] Jeffrey O Kephart and David M Chess. The vision of autonomic computing. *Computer*, 36(1):41–50, 2003.

[9] Niko Kiukkonen, Jan Blom, Olivier Dousse, Daniel Gatica-Perez, and Juha Laurila. Towards rich mobile phone datasets: Lausanne data collection campaign. *Proc. ICPS, Berlin*, 68:7, 2010.

[10] Stephen Marsland. *Machine learning: an algorithmic perspective*. Chapman and Hall/CRC, 2011.

[11] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.

[12] Vincent Primault, Antoine Boutet, Sonia Ben Mokhtar, and Lionel Brunie. The long road to computational location privacy: A survey. *IEEE Communications Surveys & Tutorials*, 2018.

[13] James Blake Rawlings, David Q Mayne, and Moritz Diehl. *Model predictive control: theory, computation, and design*, volume 2. Nob Hill Publishing Madison, WI, 2017.

[14] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. Mining interesting locations and travel sequences from gps trajectories. In *WWW*, pages 791–800, 2009.