Privacy Enhancing Technologies for Ridesharing

Ulrich Matchi Aïvodji

Privacy Enhancing Technologies for Ridesharing

Ulrich Matchi Aïvodji
LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France
Email: umaivodj@laas.fr

***Abstract*—The ubiquitous world in which we live has fostered the development of technologies that take into account the context in which users are using them to deliver high quality services. For example, by providing personalized services based on the positions of users, *Location-based services* (LBS) encourage the emergence of ridesharing services. This success has come at the cost of user privacy. Indeed in current ridesharing services, users are not in control of their own location data and have to trust the ridesharing operators with the management of their data. In this paper we aim at developing privacy preserving ridesharing systems. Our first experiments proved that privacy could be improved without scarifying the utility of the delivered service.**

I. Introduction

Nowadays, massive usage of cars takes a big part in pollution and congestion observed in urban areas. Many academic works, ranged from Transportation [26], [13] and Economy [9] to Sociology [27], have agreed on the fact that dynamic ridesharing system can be of a great help to address the shortcomings of current transportation systems. Ridesharing services such as *Blablacar, Carma, Uber...* have become in many ways a good mobility alternative for users in their daily life.

Despite the usefulness of these systems, their current implementations are still facing the challenge of privacy. As mentioned in [20], privacy is an important aspect to consider when implementing ridesharing systems because of the risk of privacy loss to which ridesharing's agencies expose users while collecting their private information (*e.g.,* positions and identities) in order to deliver high quality service. For instance, a malicious person, hereafter referred to as the *adversary*, can use the location data to cause a privacy breach [21]. In particular, this adversary can learn the Points Of Interests (POIs) of ridesharing users, compute their mobility models to infer their future movements or even de-anonymize them in another location dataset [22].

In the light of all these potential risks, we follow the *privacy-by-design* principle [12] to design a privacy preserving ridesharing system.

II. Privacy Enhancing Technologies

Privacy mechanisms can be viewed as the ability of individuals to control when, how, and to what extent their personal data is available to others. In [17] privacy methods are classified in two main categories: *law-based* and *technique-based* approaches. Although law-based approaches can deter an adversary from committing a privacy breach due to the threat of heavy penalties, their actions are really visible *a posteriori*, when the offense (*i.e.*, privacy violation) is already committed. In the other hand, technique-based approaches hereafter referred to as *privacy enhancing technologies* (PETs), help in reducing the success rate of the adversary when performing an attack by the mean of several techniques. PETs are defined in [29] as a system of ICT measures protecting informational privacy by eliminating or minimising personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system. When it comes to users' spatiotemporal information, *location privacy* is defined by [5] as preventing an unwanted entity to learn the past, present and future geographic position of an individual. The main objective of location privacy techniques is to prevent the adversary from learning the mobility traces of users. We summarize in Table I the most commonly available techniques. Although these techniques are quite generic, the specificities of the service considered impact the choice of the method to adopt.

Method	**Description**
Aggregation	Aggregate mobility traces over time and space to reduce individual identification.[16], [11]
Anonymity	k-anonymity requires that each equivalence class (*i.e.*, a set of traces that are indistinguishable from each other) contains at least k records. [28], [14]
Cloaking	Blur the mobility traces of a user into a region with poor spatiotemporal resolution. [23], [15]
Encryption	Encrypt the mobility traces of a user before disclosing them. [30], [7]
Geographical masking	Perturb the original location through the addition of noise. [3], [24]
Mix zone	Mix-zones are regions wherein the locations of users are not recorded. In addition, the pseudonym of a user entering this zone differs from the one that he will have when he exits. The overall objective is to increase the unlinkability of the user.[5], [19]
Pseudonyms	Replace the identifier of a user with a pseudonym. [6], [10]

TABLE I: Overview of location privacy approaches.

The privacy-by-design principle [12] recommends to integrate privacy early in the design phase of the system (*e.g.*, system structure, hardware design, data processing, applications, etc.) instead of adding a posteriori mechanisms to enhance privacy once the system is already built and deployed. With that method in mind, we have integrated existing privacy enhancing technologies and multimodal routing algorithms to privately compute interesting meeting points for ridesharing.

III. Meeting Points in Ridesharing: a Privacy Preserving Approach

To summarize, this work is about how to solve the dynamic ridesharing problem in such a way that users do not have to disclose their origin and destination location information but

still learning the pick-up and drop-off location that will make their trip as cost effective as possible. This very first work falls into the well known research area of Secure Multiparty Computation. This paper gives an overview of our approach. More details can be found in [2].

A. Problem statement

A dynamic ridesharing scenario can be summarized as follows. We consider a driver and a rider, each having their own origin and a destination (see Figure 1). The driver is looking for a rider to pick-up and is willing to make a small detour while the rider is looking for an itinerary in which he may use ridesharing as part of his journey. The participants expect to get a response to their request as soon as possible. The objective of the routing component of the ridesharing platform is to find for both of them pick-up and drop-off locations and optimal itineraries for their journeys. The optimization function considered is the minimization of the arrival times of both users (*i.e.*, the sum of their arrival times).

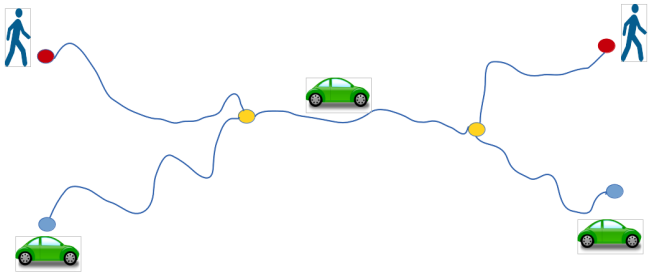

Fig. 1: Meeting points for two users in ridesharing systems.

To find ideal meeting point for the scenario described above, authors in [8] have introduced the 2 Synchronization Points Shortest Path problem (`2SP-SP`) and proposed a polynomial approach to solve it. Given an instance of the scenario, the proposed approach is able to find pick-up and drop-off locations that optimize the ridesharing cost. The privacy-preserving ridesharing problem, hereafter referred to as `Priv-2SP-SP`, can be considered as an instance of `2SP-SP` with privacy constraints in addition to existing optimization constraints. `Priv-2SP-SP` should allow us to produce solutions as good as `2SP-SP`'s ones but without having to disclose location information of any participant to a centralized third party.

B. Proposed Approach

We will now describe the core of our approach that combines secure multiparty computation and multimodal routing to implement a ridesharing service in a privacy-preserving manner.

Secure Multiparty Computation (SMC) is a branch of cryptography that aims at computing a function depending on the inputs of several parties in a distributed manner, so that only the result of the computation is revealed while the inputs of each party remain secret [25]. We particularly rely on *Private Set Intersection* (PSI) [18] that can help two parties in jointly computing respectively the intersection of their private input sets without leaking any additional information. PSI rely on Homomorphic encryption, a form of encryption that allows computations to be carried out on ciphertext, thus generating an encrypted result which, when decrypted, matches the result of operations performed on the plaintext. The main idea of PSI is to represent a set as a polynomial, the elements of the set as its roots and evaluate this polynomial homomorphically.

Multimodal routing algorithms compute shortest path on a multimodal network with several transportation modes (*e.g.*, walk, bike, car, public transportation ...). The computation of multimodal shortest paths rely on the *Regular Language Constrained Shortest Path Problem* [4] that uses a regular language L to model constraints on transportation modes. We particularly rely on isochrones to compute potential meeting points. Given an origin vertex s and a radius r, an *isochrone* is the set of vertices in the shortest-path-tree T rooted at vertex s such that the path distance from root s to any other vertex $v \in T$ is less or equal to r. Example in Figure 2 shows 7 isochrones of radius ranged from 10 to 60 minutes for a rider starting his journey at our laboratory LAAS-CNRS located at $(43.563725, 1.476744)$ in Toulouse.

Rather than working in a centralized way as in `2SP-SP`, `Priv-2SP-SP` is a decentralized approach and consists in three main steps: (1) local computation of potential meeting points (2) secure computation of shared potential pick-up and drop-off sets and (3) selection of the best pick-up and drop-off points.

We assume that the participants in the decentralized ridesharing system may be *honest-but-curious* adversaries in the cryptographic sense. In practice, this means that they will follow the recipe of the established protocol while trying to infer additional information from the output of the computation, its intermediary results and the cyphered inputs of other participants. We also assume that the two users have been put in contact by an external third party and communicate over a secure channel and have sufficient computing resources on their own platforms (*e.g.*, personal computer or smart-phone) to perform the tasks requiring local computations such as the cryptographic ones.

Step 1: Local computation of potential meeting points In this step each user locally computes its potential meeting points. More precisely, from its origin (resp. destination) location he/she infers potential pick-up (resp. drop-off) location by computing isochrones starting respectively from origin and destination locations. Any node of the transportation network that falls into an isochrone is marked as potential meeting points with its associated cost. In the set of potential pick-up locations the cost associated to each node represents the estimated time spent by the user to reach it from its origin. Similarly, in the set of potential drop-off locations the cost associated to each node represents the estimated time spent by the user to reach its destination from the selected node. To summarize, the rider (resp. driver) computes L^r_{up} and L^r_{off} (resp. L^d_{up} and L^d_{off}) where L^u_{up} (resp. L^u_{off}) denotes the potential pick-up (resp. drop-off) locations of user u.

Step 2: Secure computation of shared meeting points In this step, based on the potential meeting points the two participants compute in the previous step, they obtain shared meeting points. This is achieved using PSI while the private input set is a set of integers representing the identifier of

potential pick-up and drop-off locations. To avoid triangulation attack, cost information related to each node are kept private as this information may help in inferring users' origin and destination locations. To summarize this step securely leads to the computation of $L_{up} = L^d_{up} \cap L^r_{up}$ and $L_{off} = L^d_{off} \cap L^r_{off}$.

Step 3: Selection of ideal pick-up and drop-off points In this final step, users first compute all shared paths $\pi_{i,j}$ between points i and j where $i \in L_{up}$ and $j \in L_{off}$. Then based on the private cost of common meeting points and the traveling cost associated to each shared path, each user computes and assigns a score ($sc^d(\pi_{i,j})$ for the driver and $sc^r(\pi_{i,j})$ for the rider) to each shared path which reflects his/her willingness to use the selected shared path in his/her trip so that the path with the highest score will be the most mutually interesting for both driver and rider. Computed scores are then exchanged so that each participant knows the score given by the other party to any shared path $\pi_{i,j}$ along with the score he/she has assigned to it. Finally the elected path will be the one of the highest cumulative score. To summarize, this step returns the optimal pair of pick-up and drop-off locations $(\mathrm{i}^*, \mathrm{j}^*) = \mathrm{argmax}_{(i,j)} \ (sc^d(\pi_{i,j}) + sc^r(\pi_{i,j}))$.

By following the steps mentioned above, users are now able to find optimal meeting points *w.r.t.* privacy constraints for ridesharing.

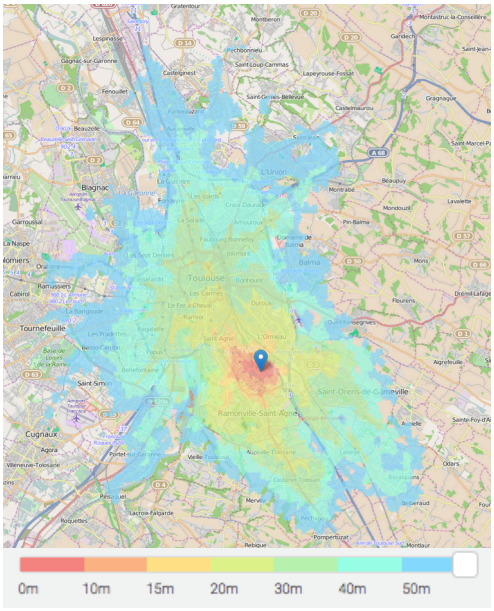

Fig. 2: Examples of isochrones.

C. Security and performance analysis

Security analysis.
Since we solve `Priv-2SP-SP` by using a PSI method, the security of our scheme depends mainly on the security of the PSI method used. In [18], the authors proved the security of PSI in the semi-honest model.

Communication and computational complexities.
The complexity of the proposed approach is directly proportional to the complexity of the shortest path algorithm. More precisely, each participant runs 2 isochrones algorithms to get their potential pick-up and drop-off locations. Finding the common POIs using PSI has a linear complexity. Once common pick-up and drop-off locations are discovered, the participants run $|V|$ shortest-path algorithms in the worst case to get all the $|V|^2$ possible shared paths. The scoring and ranking step requires the insertion of $|V|^2$ paths into a binary heap. All these steps lead to a global complexity of $\mathcal{O}(|V| \times |E| \times \log |V|)$. Finally, the communication complexity is $\mathcal{O}(|V| \times \log |V|)$ as the identifier of each node may have to be exchanged after being encrypted. The respective worst case complexities of the exact and centralized approach (`2SP-SP`) and the privacy-preserving approach (`Priv-2SP-SP`) are summarized in Table II.

	Communication cost	**Computational cost**
`2SP-SP`	$\mathcal{O}(1)$	$\mathcal{O}(\|E\| \times \|V\|^2)$
`Priv-2SP-SP`	$\mathcal{O}(\|V\| \times \log \|V\|)$	$\mathcal{O}(\|E\| \times \|V\| \times \log \|V\|)$

TABLE II: Cost and runtime analysis.

D. Evaluation

To validate our approach experiments have been performed with 100 randomly generated instances of the ridesharing problem using a multimodal transportation graph of the city of Toulouse in which we consider any vertex as potential ridesharing location. The multimodal graph has been generated using Openstreetmap's[1] data for the road network and Tisseo's GTFS[2] data for the public transportation network. The resulting graph has the following characteristics: $|V| = 75837$, $|E| = 527053$, $Modes = \{\text{Bus, Walk, Car, Subway, Tramway}\}$.

The multimodal routing algorithms have been implemented in C++ as well as the cryptographic primitives using the NFLlib library [1]. Experiments were run on a virtual machine with 5GB RAM on a 2,9 GHz Intel Core i7 host machine.

We then compared the performances of our privacy preserving approach to `2SP-SP` from both ridesharing cost and runtime point of view. Results of our experiments (average values and standard deviations) are reported in Table III.

	Cost (s)		Runtime (s)	
	Mean	SD	Mean	SD
`2SP-SP`	2872.03	672.01	0.75	0.55
`Priv-2SP-SP`	2941.25	679.20	0.48	0.30

TABLE III: Cost and runtime analysis.

To summarize, *w.r.t.* the runtime `Priv-2SP-SP` is slightly better than `2SP-SP`. This confirms our complexity analysis and has been possible thanks to the optimization of the shared path computation and the use of modern cryptographic tools in PSI [1]. As far as the quality of the solution is concerned (*i.e,* ridesharing cost) the average gap between our solutions and the conventional `2SP-SP` solutions is of 2.41%.

[1] http://www.openstreetmap.org
[2] https://developers.google.com/transit/gtfs

E. Scalability

The computational complexity of `Priv-2SP-SP` can be expressed more generally as $\mathcal{O}(|E|\times|N|\times\log|V|)$ where N represents the set of potential ridesharing locations. We consider in our experiments the worse case in which $N = V$. The reduction of N to a set of user-defined ridesharing locations, highly desirable in practice, will make the algorithm more scalable as both the runtime required by the computation of shared paths and the runtime of the private set intersection grow with N.

IV. Conclusion

In this paper, we have proposed a distributed and secure algorithm enabling users of ridesharing services to interact in a private manner. We rely on local computations, secure multiparty computation and multimodal routing algorithms, to obtain a ridesharing itinerary for users. Our results and analysis show that privacy enhancing technologies can help in solving the ridesharing problem while respecting the privacy of users and obtaining results qualitatively comparable to the one obtained from centralized infrastructures. For the ridesharing application, the proposed privacy preserving approach does not greatly impact the quality of the solutions with regard to optimal ones, moreover, it leads to a lower run time.

Our future work will investigate the generic version of ridesharing problem, involving multiple drivers and multiple riders, and study the corresponding privacy-preserving matching problem. Finally in a broader context, we will also explore how secure multiparty computation techniques can be used to solve other mobility problems. For instance in geosocial network, one could be interested in finding common meeting points for a set of users by combining multimodal routing algorithm with their social preferences.

References

[1] Carlos Aguilar-Melchor, Joris Barrier, Serge Guelton, Adrien Guinet, Marc-Olivier Killijian, and Tancrede Lepoint. Nfllib: Ntt-based fast lattice library. In *RSA Conference Cryptographers' Track*, 2016.

[2] Ulrich Matchi Aïvodji, Sébastien Gambs, Marie-José Huguet, and Marc-Olivier Killijian. Meeting points in ridesharing: a privacy-preserving approach. *LAAS-CNRS Report*, 2016.

[3] Marc P. Armstrong, Gerard Rushton, and Dale L. Zimmerman. Geographically masking health data to preserve confidentiality. *Statistics in Medicine*, 18(5):497–525, 1999.

[4] Chris Barrett, Riko Jacob, and Madhav Marathe. Formal-language-constrained path problems. *SIAM Journal on Computing*, 30(3):809–837, 2000.

[5] Alastair R. Beresford and Franck Stajano. Location privacy in pervasive computing. *Pervasive Computing*, 2(1):46–55, 2003.

[6] Claudio Bettini, Sergio Mascetti, X Sean Wang, Dario Freni, and Sushil Jajodia. Anonymity and historical-anonymity in location-based services. In *Privacy in Location-Based Applications*, pages 1–30. Springer, 2009.

[7] Igor Bilogrevic, Murtuza Jadliwala, Vishal Joneja, Kubra Kalkan, Jean-Pierre Hubaux, and Imad Aad. Privacy-preserving optimal meeting location determination on mobile devices. *Information Forensics and Security, IEEE Transactions on*, 9(7):1141–1156, 2014.

[8] Arthur Bit-Monnot, Christian Artigues, Marie-José Huguet, and Marc-Olivier Killijian. Carpooling: the 2 synchronization points shortest paths problem. In *13th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS)*, pages 12–p, 2013.

[9] David Brownstone and Thomas F Golob. The effectiveness of ridesharing incentives: Discrete-choice models of commuting in southern california. *Regional Science and Urban Economics*, 22(1):5–24, 1992.

[10] Levente Buttyán, Tamás Holczer, and István Vajda. On the effectiveness of changing pseudonyms to provide location privacy in vanets. In *Security and Privacy in Ad-hoc and Sensor Networks*, pages 129–141. Springer, 2007.

[11] Claude Castelluccia. Securing very dynamic groups and data aggregation in wireless sensor networks. In *Mobile Adhoc and Sensor Systems, 2007. MASS 2007. IEEE International Conference on*, pages 1–9. IEEE, 2007.

[12] Ann Cavoukian. Privacy by design... take the challenge. information and privacy commissioner of ontario (canada), 2009.

[13] Nelson D Chan and Susan A Shaheen. Ridesharing in north america: Past, present, and future. *Transport Reviews*, 32(1):93–112, 2012.

[14] Rui Chen, Benjamin CM Fung, Noman Mohammed, Bipin C Desai, and Ke Wang. Privacy-preserving trajectory data publishing by local suppression. *Information Sciences*, 231:83–97, 2013.

[15] Reynold Cheng, Yu Zhang, Elisa Bertino, and Sunil Prabhakar. Preserving user location privacy in mobile data management infrastructures. In *Privacy Enhancing Technologies*, pages 393–412. Springer, 2006.

[16] Norman H. Cohen, Apratim Purakayastha, John Turek, Luke Wong, and Danny Yeh. Challenges in flexible aggregation of pervasive data. *IBM Research Division*, 1, 2001.

[17] Caitlin D. Cottrill. Approaches to privacy preservation in intelligent transportation systems and vehicle-infrastructure integration initiative. *Transportation Research Record: Journal of the Transportation Research Board*, 2129:9–15, 2009.

[18] Michael J Freedman, Kobbi Nissim, and Benny Pinkas. Efficient private matching and set intersection. In *Advances in Cryptology-EUROCRYPT 2004*, pages 1–19. Springer, 2004.

[19] Julien Freudiger, Maxim Raya, Márk Félegyházi, Panos Papadimitratos, et al. Mix-zones for location privacy in vehicular networks. In *ACM Workshop on Wireless Networking for Intelligent Transportation Systems (WiN-ITS)*, 2007.

[20] Masabumi Furuhata, Maged Dessouky, Fernando Ordóñez, Marc-Etienne Brunet, Xiaoqing Wang, and Sven Koenig. Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological*, 57:28–46, 2013.

[21] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Show me how you move and I will tell you who you are. *Transactions on Data Privacy*, 4(2):103–126, 2011.

[22] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. De-anonymization attack on geolocated data. *Journal of Computer and System Sciences*, 80(8):1597–1614, 2014.

[23] Bugra Gedik and Ling Liu. Mobieyes: A distributed location monitoring service using moving location queries. *Mobile Computing, IEEE Transactions on*, 5(10):1384–1402, 2006.

[24] Mei-Po Kwan, Irene Casas, and Ben C. Schmitz. Protection of geoprivacy and accuracy of spatial information : How effective are geographical masks? *Cartographica : The International Journal for Geographic Information and Geovisualization*, 39(2):15– 28, 2004.

[25] Yehuda Lindell and Benny Pinkas. Secure multiparty computation for privacy-preserving data mining. *Journal of Privacy and Confidentiality*, 1(1):5, 2009.

[26] Catherine Morency. The ambivalence of ridesharing. *Transportation*, 34(2):239–253, 2007.

[27] Annika M Nordlund and Jörgen Garvill. Effects of values, problem awareness, and personal norm on willingness to reduce personal car use. *Journal of environmental psychology*, 23(4):339–347, 2003.

[28] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[29] GW Van Blarkom, JJ Borking, and JGE Olk. Handbook of privacy and privacy-enhancing technologies. *Privacy Incorporated Software Agent (PISA) Consortium, The Hague*, 2003.

[30] Yong Xi, Loren Schwiebert, and Weisong Shi. Privacy preserving shortest path routing with an application to navigation. *Pervasive and Mobile Computing*, 13:142 – 149, 2014.