Towards Privacy-Aware Location-Based Recommender Systems

Marcello Paolo Scipioni

University of Lugano (USI), Faculty of Informatics
Via G. Buffi, 13, 6904 Lugano, Switzerland
marcello.paolo.scipioni@usi.ch

Abstract. With the diffusion of new generations of GPS-enabled smart-phones, location-based recommender systems are appearing on market, allowing users to get personalized suggestions on-the-go about new places to visit based on user tastes. Although these systems are getting very popular and their functionalities look very attractive to users, a number of location-privacy issues are arising. In this paper we give an overview of the working principles of current location-based recommender systems, analysing them from the user privacy perspective. We then call for a new approach to location-based recommendation based on homomorphic encryption in order to enhance users' privacy.

1 Introduction

On-line social networking services (SNSs) have seen a very rapid growth in the last few years. Millions of people started using this kind of services for sharing pictures and videos, as well as for chatting with friends and exchanging impressions, suggestions, or just for communicating what they do and what they like. The growing connectivity capabilities of modern smart-phones, together with the widespread diffusion of web and multimedia services, has boosted the omnipresence of SNSs [8]: people use them on-the-go, for sharing social activities and experiences, but also for acquiring information about nearby shops or events. In this rapidly changing context, several location-based mobile services have recently been introduced: Google, for instance, has developed Latitude[1], a friend-finder application, while Facebook has recently launched Places[2], an extension that allows to share user's location with friends, check for nearby friends and discover places in the vicinity [17].

Among all the location-based services, however, we focus on a specific category of applications which looks very promising in the near future: location-based recommender systems. Users can get recommendations of places like restaurants, pubs or discos on-the-go, based on user's current position and on user's tastes, as manifested by past movements and subsequent visits to different spots ("people that usually frequent this restaurant usually like to go to that other club").

[1] http://www.google.com/latitude
[2] http://www.facebook.com/places

Examples of such services are Foursquare[3] and Brightkite[4] – SNSs where people can write text messages and share them with people in their vicinity and see who has been there before – or Loopt[5] – a location-based recommender system where people can find nearby places together with user-submitted recommendations, and get suggestions about places that may be of interest based on user preferences.

Personalized location recommendation services offer very attractive features, but on the other hand novel challenges on the privacy side have to be faced when the recommendation algorithms are applied to location data. The suggestions provided by the system have to match user tastes, therefore the recommender algorithms need to personalize their suggestions accordingly; however, there is a considerable difference when we switch from generic items like books or movies to location data: the knowledge of location data tracks over long time intervals and their analysis and detailed profiling – which would naturally be performed in state-of-the-art centralized approach where the service is aware of all clients data – can be strongly invasive of users' personal privacy.

In this paper, we focus on location-based recommender systems and on the arising privacy issues related to them. A brief insight into recommender systems and recommendation algorithms is given to highlight their general working principles; we then focus on location-based recommendation and on the peculiarities of location data in this context, especially from a privacy perspective. Finally, we call for a new privacy-aware approach to perform location-based recommendation based on homomorphic encryption. The goal of this preliminary work is to analyse location-based recommender systems and highlight the arising issues on the privacy side, and to identify the challenges to be addressed in location-based recommender systems being thus able to formulate new privacy-aware approaches to the problem.

2 Moving to Location-Based Recommendations

In a recommender system users are suggested new items based on their tastes; for instance, the popular on-line bookstore Amazon suggests new books to buy based on the previous purchases of a certain user [13]. In each recommender system some knowledge about the user, needed to match his tastes, is collected in a *"user profile"*; at the same time, each item is described by means of characteristic attributes (e.g. author, subject or title of a book), which contribute in forming an item profile, called *"content"*. Each item can be evaluated by users through *"ratings"*, which represent the utility of an item to a user [1]. The problem of recommendation can be formulated as a rating estimation problem: the unknown item ratings for a certain user who asks for a recommendation – the *active user* – need to be estimated by the system; the k items with the highest estimated

[3] `http://foursquare.com`
[4] `http://brightkite.com`
[5] `http://www.loopt.com`

ratings are then chosen for recommendation. The rating estimation can be addressed either by considering items liked by *other users* with similar tastes, or by considering items similar to the ones *the same user* likes. In the former case we have *"collaborative recommendations"*, while in the latter one *"content-based recommendations"*; hybrid approaches combine both of them together [1]. Several state-of-the-art recommendation algorithms rely on collaborative filtering techniques, where the similarity among users is computed (e.g. through Pearson correlation coefficient) and the ratings are estimated from the known ratings of other users and weighted according to the similarity measure considered [9].

In location-based recommender systems, items are represented by *"places"*: users may be interested in restaurants, shops, clubs, cinemas as well as in concerts, events or tourist attractions. Such systems not only need to provide effective recommendations for users, but should also be designed to be used in mobile settings; for this reason a number of constraints and need to be taken into account while designing such systems, due to their inherent mobile nature. The general theory about recommender systems, however, is not straightforwardly applicable, but needs to be adapted for the following reasons:

- user profiles need to be derived starting from raw location data originated by user's smartphones;
- characteristic attributes for places need to be retrieved from external sources;
- ratings may not be explicitly expressed by users.

To derive a user profile, location data need to be processed and clustered into stay-points characterized by given space boundaries, where users have stayed longer than a given time threshold, and connected together by edges modelling the temporal sequence of visited stay-points in a graph structure [18]. The stay-points characterizing a user profile, however, should be meaningful spots for the user (e.g. shops, cinemas, clubs), and not just spaces where the user happens to be (e.g. stuck in the city traffic): we will refer to them as *"places"*, opposed to *"spaces"*, using the terminology suggested by Mancini et al. [14].

Many positioning technologies are now available on mobile platforms: GPS, cell-id based mobile network positioning and Wi-Fi fingerprinting[6] not only provide location at different granularity, but also have different privacy implications. GPS is a self-positioning method, which has the advantage of being independently computable on chip, without any interaction with external parties; it is however more battery-draining than other methods [11]. Moreover, having only GPS data may be insufficient to identify the right place visited by a user: in indoor conditions or in dense urban areas, where GPS quality is strongly degraded, it may be hard to determine whether a certain user in a shopping mall is e.g. at a Chinese restaurant or in the shoe-shop right next to it. The ability to distinguish meaningful places might be enhanced by allowing users to *"check-in"* explicitly into available venues in the vicinity (explicit check-ins are strongly used in location-based social networks). Other positioning technologies

[6] For cell-id and Wi-Fi-based positioning, see, e.g. `http://www.skyhookwireless.com`

can be exploited to distinguish places: network-based positioning might be helpful in urban areas, where mobile phone cells are more densely distributed; it is though necessary to interact with an external entity (which might be a service or a network provider) to resolve the cell-ids to geographic coordinates. Wi-Fi fingerprinting works best indoors, and might represent the best solution for the identification of places, when available. Wi-Fi fingerprinting and cell-id positioning are remote positioning technologies dependent from external service providers, which is a drawback from the privacy side because location data are exposed to external entities, but with a careful cashing of past data these threats can be mitigated.

A location-based recommender system might not account for explicit ratings expressed by users; in this case, implicit ratings can be inferred by the system. Some places might appear multiple times in the list of frequented places, and frequent visits to the same place or places where the user spends long periods of time can be considered "liked" by the user. This assumption would however not hold for places that the user visited just once, and no assumptions can be made anyway with respect to places "disliked" by the user. To solve this issue, allowing users to explicitly rate their liked/disliked places can increase the quality of recommendations, but would require a certain effort from the user, who may feel bothered by frequent feedback requests. The most common recommendation algorithms, e.g. collaborative filtering, use explicit ratings; it is worth exploring more in detail the adoption of implicit ratings in recommender system research.

Another issue arises from the nature of items: since we are dealing with places, which can refer to shops (subject to ownership changes), events (which might be created by users and are limited in time) and, more in general, with venues which can vary over time – and not with a predefined set of products periodically revised by the producer company – the system must keep track of the validity of item data over time. The Google Places API[7] offers a very useful interface for incorporating existing places; however, the system must also allow users to create new venues and, possibly, check their validity.

Finally, another interesting feature of location-based recommender systems is the possibility to ask for suggestions for the *next* place to visit, taking care not only of a single place but considering sequences of places. This is an interesting development which may naturally follow in the research on location-based recommender systems.

3 Location Privacy in Recommender Systems

The core functionality of location-based recommender systems, which consists of suggesting places to users, differs considerably in the communication patterns from that offered by location-based social networks. While in social networks the natural information stream follows the interpersonal relationship among friends, in recommender systems the information contained in a user profile is combined

[7] See `http://code.google.com/apis/maps/documentation/places`

by the system together with the tastes of others and proposed back in form of recommendations to match the profile of the requesting user [16]. The matching phase among user profiles through which the provider offers personalized suggestions is the core functionality of the service; for this reason, user profiles represent one of the essential features in a recommender system.

However, the disclosure of user's profiles allows to uncover many facets of people's life, raising many privacy issues [3]. More specifically, the automatic inspection of past user location data can be a big threat to personal privacy [12]. Moreover, other privacy risks are given by the possible reuse of previously compiled user profiles at later time, e.g. for advertising reasons (unsolicited marketing), or by the possibility to sell profile data to third parties.

The simple pseudonymization of all location tracks would not be sufficient for preventing later identification of the users by the service provider. Hoh et al. [10] recorded GPS data from the vehicles of 65 people, through which they succeeded in identifying 85% of people's home locations from GPS tracks with 1 minute sampling. At lower time intervals (4 - 10 minutes) they could still identify around 40% of correct home locations. As also Gruteser and Grunwald point out, restricted space identification attacks can be used to identify people based on (even anonymous) location data combined with publicly available data [7]. Obfuscation-based techniques which degrade the spatio-temporal accuracy of location data, like those proposed by Gruteser and Grunwald [6], would not be applicable in this case, since the precision of the location data is a key feature in order to provide effective recommendations; degraded quality of location data would cause the system not to work properly, compromising its functionality.

The solution we propose to overcome the privacy problem is to keep all user data private. All user data should be stored on the user's smart-phone, and the clustering of location data into places and the creation of user profiles should also be performed on the user's side. All user profile data would be kept private, and the property of these data would remain with the users who originated them.

Ratings of places, either explicitly expressed by users or implicitly inferred by the system, hold all the available knowledge about users. All the known ratings, together with the item profiles, are used to estimate the ratings of items unknown to the active user; the items with the highest ratings are then proposed as a recommendation. For the importance that ratings hold, we propose to store them in encrypted form, using homomorphic encryption. Homomorphic encryption is a particular category of encryption functions where the computation of given algebraic operations on the ciphertext (without the need of decrypting it) is equivalent to the computation of other operations on the plaintext [15]. This allows to perform calculations on the encrypted data, without disclosing any private data. More formally, homomorphic encryption can be defined, according to [5], as:

$$\forall m_1, m_2 \in \mathcal{M}, \qquad E(m_1 \odot_{\mathcal{M}} m_2) \longleftarrow E(m_1) \odot_{\mathcal{C}} E(m_2)$$

where respectively $\mathcal{M}$ represents the plaintext domain, $\mathcal{C}$ the ciphertext domain, $\odot_{\mathcal{M}}$ an algebraic operation in the plaintext domain and $\odot_{\mathcal{C}}$ the corresponding algebraic operation in the ciphertext domain.

Canny [2] proposed a protocol for collaborative filtering based on homomorphic encryption, where the vectors containing user ratings are used to compute a global model of user preferences through Singular Value Decomposition (SVD), but are never disclosed to others. All data – both the global model and the vectors of user ratings – are encrypted with the ElGamal cryptosystem [4]. The model is computed from partial results submitted by the clients and combined together through a scheme which exploits the homomorphic properties of the ElGamal cryptosystem; moreover, the correctness of the partial results submitted by each client are checked through zero knowledge proofs, to avoid malicious behaviour. All quantities are encrypted using a public key, corresponding to a private key which nobody holds, but which is shared among all the clients and which can only be recovered from partial key shares of at least tn clients, where t is the fraction of total users n which are at least needed for decryption in a threshold decryption system like this.

The algorithm proposed by Canny can be applied to location based recommendation for a privacy-aware solution of the problem. The advantages of this approach look promising: users would use ratings in encrypted form, and all the estimate ratings and recommendations would be computed using encrypted data. The community of users would be able to decrypt the computed model which could be used locally for recommendations. Furthermore, Canny's protocol guarantees that no data belonging to any single users can be recovered from the decrypted model [2].

Although employing this protocol seems very attractive, its adaptation for use in the context of location-aware recommendation is not straightforward. The threshold decryption system has the advantage of preventing any single user from decrypting the whole model by himself. The threshold mechanism, however, also implies that groups of users share their partial knowledge together to be able to decrypt the model; the design of this sharing part has to take into account actual usage scenarios of the system, letting users understand its principle at the same time. This is a challenging issue that will need to be faced in the next stages of this research. Also the choice of the threshold parameter[8] should be taken balancing the security requisites with its feasibility within the implemented system: a threshold value too large would require that users share their partial knowledge with a too large subset of total users of the system.

Moreover, using features like items based on places and implicit ratings in recommender system algorithms is challenging. A careful choice of locating technologies will be needed to address these challenge, together with robust mechanisms for clustering location data into meaningful *places* and effective ways to add new places and/or their description.

[8] Canny proposes 0.2 for the value of the threshold t.

4 Summary and Outlook

In this paper we have briefly described the general principle of location-based recommender systems in order to highlight their working functionalities and the privacy issues involved in the process. Our suggested approach goes in the direction of creating a new generation of location-based recommender systems to provide recommendations without the need of collecting in a centralized way and inspecting huge quantities of data stored in the clear into user profiles.

Much work still remains to be done on this topic. In the next steps of this research we will focus on the presented challenges to apply Canny's scheme on location-based items and implicit ratings described here. The features and the performances of the resulting work will have to be compared with those of existing systems. We believe, however, that users should be given the chance of choosing recommender system with a higher level of privacy than what it is today. With a growing interest in mobile applications and personalized recommendations, the relevance of privacy-aware location-based recommender systems will likely increase in the near future.

Acknowledgments. This work is being done within the PALS project funded by the Swiss National Science Foundation under grant n. 200021_129674.

References

1. G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749, December 2005.
2. J. Canny. Collaborative filtering with privacy. In *Proceedings of the 2002 IEEE Symposium on Security and Privacy*, pages 45–57. IEEE, 2002.
3. L. F. Cranor. 'I didn't buy it for myself' privacy and ecommerce personalization. In *Proceedings of the 2003 ACM workshop on Privacy in the electronic society*, WPES '03, pages 111–117, New York, USA, 2003. ACM.
4. T. ElGamal. A public key cryptosystem and a signature scheme based on discrete logarithms. In *Advances in Cryptology*, pages 10–18. Springer, 1985.
5. C. Fontaine and F. Galand. A survey of homomorphic encryption for nonspecialists. *"EURASIP Journal on Information Security"*, 2007:1–15, 2007.
6. M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services (MobiSys '03)*, pages 31–42, New York, USA, 2003. ACM Press.
7. M. Gruteser and D. Grunwald. A methodological assessment of location privacy risks in wireless hotspot networks. In *In Proceedings of the First International Conference on Security in Pervasive Computing*, pages 10–24. Springer, 2003.
8. M. Hamblen. Smartphones, other mobile devices boost social networks, September 2010. `http://www.computerworld.com/s/article/9185038/Smartphones_other_mobile_devices_boost_social_networks`.

9. J. Herlocker, J. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 230–237. ACM, 1999.
10. B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Enhancing security and privacy in traffic-monitoring systems. *IEEE Pervasive Computing*, 5:38–46, 2006.
11. M. Kjaergaard. Minimizing the power consumption of location-based services on mobile phones. *IEEE Pervasive Computing*, 99(PrePrints), 2010.
12. J. Krumm. A survey of computational location privacy. *Personal Ubiquitous Comput.*, 13:391–399, August 2009.
13. G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1):76–80, 2003.
14. C. Mancini, K. Thomas, Y. Rogers, B. Price, L. Jedrzejczyk, A. Bandara, A. Joinson, and B. Nuseibeh. From spaces to places: emerging contexts in mobile privacy. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 1–10. ACM, 2009.
15. S. Rane, W. Sun, and A. Vetro. Secure distortion computation among untrusting parties using homomorphic encryption. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 1485–1488. IEEE, 2009.
16. M. P. Scipioni and M. Langheinrich. I'm Here! Privacy Challenges in Mobile Location Sharing. Second International Workshop on Security and Privacy in Spontaneous Interaction and Mobile Phone Use (IWSSI/SPMU 2010), Helsinki, Finland, May 2010. Co-located with Pervasive 2010.
17. R. Strohmeyer. Facebook checks in to location-based services with Places, August 2010. `http://www.computerworld.com/s/article/9180898/Facebook_checks_in_to_location_based_services_with_Places`.
18. Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W. Ma. Recommending friends and locations based on individual location history. *ACM Transactions on the Web (TWEB)*, 5(1):5, 2011.