Virat Shejwalkar

Application Specific Evaluation of Privacy Preserving Mechanisms

Due to growing use of the location based services (LBSs), concern about the privacy of users' location data is growing. Many Location Privacy Preserving Mechanisms (LPPMs) are proposed in the literature which can be roughly classified in two categories - location obfuscation and anonymization. Using the latter, camouflaging user's actual location with fakes is a common strategy but it requires fake location data that a user can use. Bindschaedler et al.[1] propose an LPPM which synthesizes such a fake location data using real location data of actual users as a seed. It is important to evaluate privacy and utility of such LPPMs and of the synthetic data. [1–4, 6] propose few metrics for the evaluation e.g. estimation error of adversarial inference as a privacy metric [3] and distance based quality loss as a utility metric [2]. But, these are very generic and do not consider application specific usage of the users' location data. This may render the conclusions based on aforementioned metrics imprecise. This will also help LBS developers choose one LPPM over the other based on the type of the application. Hence, it is important to devise application specific utility and privacy metrics and evaluate current LPPMs.

The Synthetic Location Trace Generator (SLTG) tool proposed by Bindschaedler et al.[1] is an LPPM that generates synthetic traces by capturing both semantic and geographic features of real location traces. The process of trace generation itself is privacy preserving and ensures plausible deniability. Plausible deniability here means - a trace generated from a seed can also be generated from one or more other seeds; this ensures that analysis of fake traces cannot be linked to any particular real seed location trace. The privacy tests applied to each synthesized trace ensure maximum semantic similarity and minimum geographic similarity; the similarities can be adjusted but these are the recommended settings. High semantic similarity reflects the fact that the probability with which humans move from one location, say home, to other, say work or the time they spend in those locations or the time of the day/week they are present in those locations does not vary. The low geographical similarity, however, should be low to preserve privacy of the seed locations used to generate the fakes.

Generated traces are tested for privacy vs utility trade-off. For utility loss, the authors in [1] argue that the user's true location will be embedded in $k-1$ fake locations thereby ensuring no quality loss. The utility loss experienced will not be in terms of reduced service quality but in terms of extra bandwidth consumption by the mobile device and the profile pollution[1]. This proposed utility measure would fail in for many applications due to peculiar ways in which location data is used. Take an example of ride-hailing service such as Uber. Embedding the true location in multiple fake locations will give Uber an impression that multiple users in an area are requesting a ride. Hence, due to surge pricing, it will increase the price for the ride. Along with user's utility, Uber's i.e. service provider's utility should also be considered. e.g. when the ride requested with multiple fakes is finished, the service provider will not be able to charge the correct user due to the lack of account information of all the requesting users. For the class of LBSs that link user's account to various payment gateways sending the fakes will not work. Revisiting the Uber example, while charging a user for her ride, if Uber cannot find accounts for the fake requests, it will know the location corresponding to real request endangering the user privacy.

Another common continuous LBS is navigation where user location is continuously reported to the server. Now, consider a health-care application that measures calories burnt by a user while jogging. In this case, the user's location privacy will be preserved using anonymization technique proposed. However, bandwidth consumption for continuous navigation should be calculated because sending multiple locations repeatedly might adversely affect cellular data usage. This applies to all the applications that fall in continuous LBS category. Bandwidth consumption should be especially measured for recommendation services as depending on the location, number of recommendations and so the consumption will change. But, the SLTG evaluation in [1] does not consider continuous LBS and evaluation is done only for location check-in services.

Chatzikokolakis et al.[2] propose a location remapping method post LPPM for improving utility of the LPPM. This is specifically for the LPPMs that use real location obfuscation. Here, after location is obfuscated, it is remapped to another location such that a particular quality loss function is minimized. The quality loss is proportional to $d_Q(x, z)$ where d_Q is the quality loss metric based on distance (Euclidean or Hamming distance) between actual (x) and obfuscated location (z). Such a generic metric cannot precisely estimate utility of an LPPM for different LBSs. For example, consider a weather and a restaurant recommendation application. Both of these use user's location to provide requisite services. However, the utility loss of the earlier is negligible compared to the latter if, say, location is obfuscated by few miles. Because, weather does not change if one moves from a place by few miles but list of nearby restaurants will substantially change. However, the distance based metric considered in the paper implies exactly same utility loss for both the services when same LPPM is used for the obfuscation.

Bilogrevic et al.[4] study motivations behind location check-ins of users and classify them in different categories e.g. inform about activity, appear cool/interesting, etc. Due to these purposes, they introduce *perceived utility* metric for location check-in based services. The metric links user's perceived expectation from the check-ins to different types of LPPMs that either use semantic or geographical obfuscation. It is finally concluded that LPPMs that perform semantic obfuscation harm the perceived utility of users' check-ins more than those which perform obfuscation on geographic level. This distinction would not be possible using previously proposed generic metrics. Hence, it is important to devise such application specific metrics so that utility and privacy conscious developers would choose LPPMs more wisely. As in this case, check-in based applications should use LPPMs that perform geographic obfuscation to preserve users' utility.

Along with the utility, privacy metrics should also need to be reconsidered with respect to the application specific requirements and usage of the users' location data. Shokri et al.[3] have proposed a systematic way of measuring privacy of different LPPMs based on different classes of inference attacks. The paper proposes a hypothesis that the user privacy is equal to incorrectness of adversary's inference. So, the privacy measured is probabilistic and measured as a distance between actual and inferred locations on a set of locations within which a user can move. If adversary infers exact correct location of the user the distance is 0, in all other cases it's 1. This privacy metrics is very coarse and application specific finer metrics need to be devised. For example, consider two LPPMs, inference attacks on which correctly detect the exact user location 50% of times and rest of the times accuracy of the attacks is more for one LPPM than the other. However, the metric used implies both LPPMs to be equally privacy preserving.

There are different privacy metrics proposed in the literature and it is important to compare these different metrics. Shokri et al. [3] perform such a comparison based on the privacy metric described before. But it should be further classified with respect to different settings in which some LPPMs work better or worse than the others. This will help in matching different applications and LPPMs that fall in a same class. Montazeri et al. [6] propose a metric based on mutual information between actual traces and the traces observed by an adversary; the LPPMs considered here are based on anonymization of users' location traces. Apart from this, there are well-known metrics like K-anonymity and entropy.

1 Summary

To identify the frontier, let us consider all that is known. For utility, it is clear from [1, 2] that changing the nature of LPPMs (between obfuscation and anonymization), utility metrics change and for privacy, current metrics are known to be strict and follow Hamming distance approach. The earlier raises an obvious question - if the nature of an application changes should the utility metric change? As the current literature does not explore application specific utility measurements, I would like to study requirements of applications and possibly classify the LBSs based on different ways in which location is used by different LBSs.

To this end, I would like to consider SLTG which implements LPPM based on anonymization. Reason for the choice is two-fold: 1) It is the only tool currently available that considers both semantic and geographical similarities and has high adversarial inference error. 2) If the data generation is perfected, it can be used for wide range of applications as well as research that lacks location data due to non-disclosure policies of LBSs like Uber.

As explained before, proposed in privacy metric in [3] is very coarse but highly adapted in subsequent literature and should be compared against other privacy metrics available in the literature, such as proposed by Montazeri et al.[6] which is based on mutual information between observed and actual location traces of a user. I would like to compare two privacy metrics and see if one is better than other under any circumstances.

All in all, the related work in privacy and utility measurements lacks application specific metrics

References

[1] V. Bindschaedler, R. Shokri, Synthesizing Plausible Privacy-Preserving Location Traces, IEEE Symposium on Security and Privacy (S&P) (2016)

[2] K. Chatzikokolakis, E. ElSalamouny, C. Palamidessi, Efficient Utility Improvement for Location Privacy, PoPETs (2017) (4)

[3] R. Shokri, G. Theodorakopoulos, J. L. Boudec, J. P. Hubaux, Quantifying Location Privacy, IEEE Symposium on Security and Privacy (S&P) (2011)

[4] I. Bilogrevic, K. Huguenin, S. Mihaila, R. Shokri, J. P. Hubaux, Predicting Users' Motivations behind Location Check-Ins and Utility Implications of Privacy Protection Mechanisms, NDSS Symposium (2015)

[5] A. Pham et al., PrivateRide: A Privacy-Enhanced Ride-Hailing Service, PoPETs (2017) (2)

[6] Z. Montazeri, A. Houmansadr,H. Pishro-Nik Achieving Perfect Location Privacy in Wireless Devices Using Anonymization, IEEE Transactions on Information Forensics and Security (2017)