Taxonomy of Dummy Generation Techniques for Preserving Location Privacy

[1] Priti Jagwani , [2] Ranjan Kumar

[1,2] Dept. of Computer Science, RLA (E) College, University of Delhi, 110021,India

Abstract - With the advent of GIS, Internet and mobile telephony, Location based services are becoming the need of the day. Privacy issues arising while using location based services have the potential to become the serious apprehensions. Existing privacy techniques based on anonymity generally fails in to offer the required privacy. We are focusing on the dummy generation techniques available in literature to protect location privacy. In these techniques, the actual location of a user along with several false position data (dummies) sent to the service provider. Because the service provider cannot distinguish the true position data, the user's location privacy is protected. All the available algorithms to generate dummies (for query, locations as well as for trajectories) are reviewed. We briefly discussed and compared all the dummy generation techniques and listed the details of each technique.

Keywords: *Location Privacy, Dummy Generation Techniques, Plausible Deniable search, query privacy, location privacy*

1. Introduction

With the rapid development of Internet, mobile Telephony and GIS, a new term has coined – location based services. The term location based services (LBS) is a class of applications that provide mobile users personalized services from their current locations using one of several positioning technologies, e.g. GPS, cell-phone positioning, and positioning through Wi-Fi access points. Examples of location based services include wireless 911 emergency service, geocoding, reverse geocoding, traffic advisories, location-aware advertising, tourist services, location-based games, and navigation. Other services which can be accomplished with the help of the above technology can combine personal preference (recommenders) information with present locations to facilitate users find food, lodging, and entertainment according to their tastes and pocketbooks.

On one side of the coin there is convenience brought by this new class of applications, but on the very other hand there is a worry about a new type of privacy threat, namely, the location privacy threat. These are issues introduced by the releasing of location information to untrusted third parties. Typically, a location based system can be formed by using one of the two architectures: Trusted third party (TTP) architecture and without trusted third party. In TTP architectures, client sends the request to TTP (middleware) and then middleware after securing the location and (or) identity of client forwards these requests to location server. This location server is the untrusted party and is actual service provider while middleware is considered as trusted one. But middleware itself is a single point of attack and is vulnerable. In another class of architectures that is without trusted third party (middleware) client himself is responsible for hiding the location. For this purpose he can take help of his peers/other users present in a specific area.

The basic idea is to save location and identity information of client from untrusted party (location server) or any other adversaries. Location information is sensitive because it is ubiquitous and can lead to much other information. For example, the frequent visit to a hospital may release a person's health condition. The risks of locational privacy breach range from releasing information about visits to sensitive places to enabling unwanted virtual or physical stalking. There are various techniques available for location privacy namely policy strategies, regulatory techniques, location obfuscation, data transformation and private information retrieval techniques. Among all these variety of techniques location obfuscation includes a wide group of methods to protect location information. Location obfuscation is the process of slightly altering, substituting or generalizing the location in order to avoid reflecting real, precise position. Various methods available under the head of location obfuscation are use of pseudonyms, spatial cloaking, adding random noise and dummies, rounding location information and redefinition of possible locations.

2. Dummy Generation Techniques Revisited

Among all the above mentioned methods dummy generation is a widely used and a promising strategy to protect location privacy. In this technique a user sends true

position data with several false position data ('dummies') to a service provider, who creates a reply message for each received position data
e.g. Assume $Lx = (X_1, Y_1)$ shows location of a user at time t. A message S from the user to request a service is of the form:

$$S = (u, L_1, L_2, \_ \_ \_ , L_k)$$

where u shows a user ID and $(L_1, L_2, \_ \_ \_ , L_k)$ shows a set of position data that includes one true position data and k-1 dummy locations. This request is sent to service provider. On the other hand, a service answer message R from the service provider to the user is of the form:

$$R = ((L_1, D_1), (L_2, D_2), \_ \_ \_ , (L_k, D_k))$$

where $(D_1, D_2, \_ \_ \_ , D_k)$ shows the reply of the service request corresponding to the locations $(L_1, L_2, \_ \_ \_ , L_k)$. Here other k-1 locations sent with the actual location of user are considered as noise data/dummies consisting of false position data. On getting the reply in the above mentioned form user filters the data required according to his true location. The user simply extracts the necessary information from the reply message. Hence user's true location is kept hidden from location server. In this manner, even if the service provider stores the set of position data, it cannot distinguish the true position data from the set of fake position data.

The responsibility of dummy generation can be delegated to middleware in trusted third party architectures while in decentralized architecture (where trusted third party is not available) this responsibility can be taken by the client himself as shown in the above example. The decentralized approach entitles many advantages such as scalability issues and elimination of trusted third party which can be a bottleneck. Moreover trust assumptions may not be realistic. Also, a mobile terminal in case of decentralized architecture does not need to report its location periodically to an anonymizer, as is needed in spatial cloaking solutions where the anonymizer needs up-to-date location information from all mobile terminals in order to do the cloaking. Following sections present a detailed overview of different dummy generation approaches available for location as well as for query and trajectories.

2.1 Location privacy by using dummies

Paper [2] proposed an anonymous communication technique to protect the location privacy of the users of location-based services. They also described an efficient dummy generation algorithm based on the property that dummies should not be distinguishable from true position data. If dummies are generated randomly, observers can easily find differences between true position data and dummies. In this case, location anonymity is reduced. To avoid this, the dummy must not behave completely different from the true position data. They presented the following two dummy generation algorithms to prevent service providers from finding the true position data. In these algorithms, the locations of the first dummies are decided randomly and then these will be refined further.

Moving in a Neighborhood (MN): In this algorithm, the next position of the dummy is decided in a neighborhood of the current position of the dummy. The communication device of the user memorizes the previous position of each dummy. Then the device generates dummies around the memory.

Moving in a Limited Neighborhood (MLN): In this algorithm, the next position of the dummy is also decided in the neighborhood of the current position of the dummy. However, the next position is limited by the density of the region. This algorithm is adaptable in cases where the communication device of the user can get the position data of other users. Further authors in [2] have tried to optimize the whole process to reduce the communication cost.

A new privacy area aware dummy (PAD) based approach, that is capable of offering privacy-region guarantees is proposed in [6]. To achieve this, PAD uses so-called dummy locations that are deliberately generated according to either a virtual grid or circle. These cover a user's actual location, and their spatial extents are controlled by the generation algorithms. This duality feature (dummy generation and control of spatial extent) makes this technique more usable than in spatial contexts with purely k anonymity based approaches. The PAD approach only requires a lightweight server-side front-end in order for it to be integrated into an existing client/server mobile service system. In addition, query results are organized according to a compact format on the server, which not only reduces communication cost, but also facilitates the result refinement on the client side.

Authors in [6] presented two dummy generation algorithms. One generates dummies based on a virtual grid covering the user location. The other generates dummies based on a virtual circle that contains the user location. These algorithms are flexible in that the dummy generation is configurable and controllable, thus offering means of controlling the location privacy of a user. This contrasts the other works where location dummies are generated totally at random. The PAD approach can be easily integrated into existing systems that employ client/server architectures. It does not require a trusted third-party component as an anonymizer, and nor does it assume that the server is trustable. In the server side, a lightweight front-end module suffices to render the approach functional. PAD incorporates techniques that

reduce both the upstream and downstream communication between client and server. In addition, the query results to be sent to clients are organized according to a compact format with respect to all dummy locations in a query. This format not only reduces the communication cost, but also facilitates the result refinement on the client side.

A hybrid approach to combine location cloaking and dummy generation approach is proposed in [11]. Cloaking is applied to blur precise location in to a region so that location privacy can be protected. On the other hand where cloaking area is large limited number of dummies are being generated to decrease the size of region.

2.2 Trajectory privacy

One form of location privacy using dummies can result in trajectory privacy. By generating dummies that move in human trajectories, [10] shows that location privacy of mobile users can be preserved. Two schemes that generate consistent movement patterns in a long run. Guided by three parameters in user specified privacy profile, namely, short term disclosure, long-term disclosure and distance deviation are proposed in [10]. The proposed schemes derive movement trajectories for dummies. Short term disclosure (SD) specifies requirement for protecting the current user location. Thus, given a set of current locations (including true and dummy locations), SD is the probability of successfully identifying the true user location.

While long term disclosure indicates the requirement for protecting the user trajectory and distance deviation is the average of distance difference among trajectories of dummies and the user. Finally using these privacy parameters dummy trajectories are generated by two schemes. One is random pattern scheme and rotation pattern scheme. Both the schemes outperform the other available methods in terms of protecting trajectory privacy.

When a mobile user must transmit his or her location to a central server, these location reports can be accompanied by false reports that, ideally, cannot be distinguished from the true one. The realism of the false reports is important, because otherwise an attacker could filter all but the real data. Paper [4] uses the database of GPS tracks from over 250 volunteer drivers, and developed probabilistic models of driving behavior and using these probabilities to generate random start and end points, random routes, random speeds, and random GPS noise. [4] applied these models to create realistic driving trips.

2. 3 Query privacy using dummies

Privacy concerns in LBS exist on two fronts: *location privacy* and *query privacy*. Location privacy is related to the disclosure and misuse of user's location information. An example of its implication is that if a user issues an LBS query from a location within hospital premises then the adversary can associate a medical condition with the user. Query privacy, on the other hand, is related to disclosure of the service attribute. For example, frequent queries for a hospital may lead the adversary to infer that the user is having health problems. Although distinct, location privacy and query privacy are closely related. In particular, disclosure of location may in turn reveal the service attribute to the adversary. Location privacy using dummies is discussed in the previous section; this section discusses query privacy and the literature available in this area in detail.

A user-centric technique named as DUMMY-Q, for query privacy protection which operates solely on the user side and does not require any trusted third party is developed in [8]. The key idea is to confuse the adversary by issuing multiple counterfeit queries with varying service attributes but the same (real) location, henceforth referred to as *dummy queries*, along with each real query issued by the user. Aim of the proposed technique is to prevent the LBS server from correlating the service attribute. Authors in [8] claimed that in case of continuous LBS scenarios effectiveness of location obfuscation using spatial generalization aided by anonymization has been abated. So a query-perturbation-based scheme that protects query privacy in continuous LBS even when user identities are revealed is proposed.

A critical requirement for dummy generation is that the dummy service attribute values must be generated in a judicious manner so as to remain consistent with the *query context* - i.e., the location where query is issued. In addition, one must insert the same (dummy) service attribute values over different snapshots of a continuous LBS query, in order to prevent the adversary from inferring the most frequent value as the real one. Other challenges are minimization of the number of inserted dummy queries because each consumes additional overhead for issuing the query and waiting for the answer. Finally, a resource challenge faced is, the limited storage and computational capacity of mobile devices, from which many LBS queries are issued and therefore privacy protection must be enforced.

Considering the above challenges of query privacy using dummies, a dummy query generation algorithm is proposed in [8]. This algorithm takes into account two inputs for generating the dummy service attribute values:

the query context, i.e., the set of service attribute values which may be issued from a given location, and the user's motion model, i.e., the set of locations the user may travel to in future snapshots of the continuous LBS query. Based on the inputs, Pool-Builder component randomly selects a set of dummy service attribute values such that, even after the exclusion of "unreasonable" dummy values according to all future snapshots of the query, the adversary still cannot compromise the real service attribute value with probability exceeding a pre-determined threshold. Hence, query privacy is guaranteed.

[9] proposed another query privacy approach accomplished with the help of dummies. In this technique dummies are generated at the middleware (Anonymization server/ trusted third party). These dummies are generated by parameters in the dummy profile. The dummy profile is a file containing a list of all mobile users in the system along with corresponding dummy user identification numbers and profile count which is the number of dummies associated with a real mobile user on the dummy profile. profileCount is initialized to be the maximum *K* value allowed in the system.

Existing privacy protection algorithms rarely pay attention to both of query privacy and location privacy. Li min et. Al in [5] proposes a novel privacy protection method which combines **K**-anonymity and **L**-diversity to protect both location privacy and query privacy. Two effective query-privacy-aware methods are introduced into the cloaking algorithm. One is the **history sharing scheme** which confuses history queries within tolerance time. Another is the **batch query scheme** which confuses real queries presented by the peers. The main idea of the method is to realize the L-diversity along with cloaking through confusing history queries or actual queries, termed as history sharing scheme and batch query scheme respectively. Moreover, L-diversity measurements rely on query entropy, rather than considering the differences in service attributes.

3. Plausible Deniable Search

A client-centered approach of *plausibly deniable search* (PDS) for web based query search has been devised in [7]., In this each user query is substituted with a standard, closely-related query intended to fetch the desired results. In addition, a set of k-1 cover queries are issued; these have characteristics similar to the standard query but on unrelated topics. The system ensures that any of these k queries will produce the same set of k queries, giving k possible topics the user could have been searching for.
A user issuing a set of queries, $S = \{Q_1, \ldots, Q_k\}$, where $Q_i \in S$ is the desired query, has k-Plausibly Deniable privacy of Q_i if

1. the user can show that any query $Q_j \in S$, would have generated the set S with equal probability,
2. all $Q_j \in S$ are on different topics, and
3. all $Q_j \in S$ are equally plausible as actual user queries

Paper [7] used a Latent Semantic Indexing (LSI) based approach to generate queries, and evaluate on the DMOZ webpage collection to show effectiveness of the proposed approach. In this work PDS is being applied for web search queries. a similar approach for location data (as well as for web data also) is carved in [1]. Both the works tried to generate convincing fakes which is actually a complex process, and there are no guarantees as to how convincing the fakes are.

4. Conclusion and Future Directions

Dummy generation techniques for privacy are well suited with both types of architectures (with or without trusted third party). We highlighted dummy generation techniques available for location data, query and trajectory data. A domain of concern is generation of judicious dummies which cannot be distinguished from the real data. Plausible deniable search can be one promising way to address this concern. Although application of PDS for position data and location queries is a future research direction. For generation of judicious dummy data, use of location semantics can also be explored.

References

[1] Chow, Richard, and Philippe Golle. "Faking contextual data for fun, profit, and privacy." Proceedings of the 8th ACM workshop on Privacy in the electronic society. ACM, 2009.

[2] Kido, Hidetoshi, Yutaka Yanagisawa, and Tetsuji Satoh. "An anonymous communication technique using dummies for location-based services." Pervasive Services, 2005. ICPS'05. Proceedings. International Conference on. IEEE, 2005.

[3] Kido, Hidetoshi, Yutaka Yanagisawa, and Tetsuji Satoh. "Protection of location privacy using dummies for location-based services." Data Engineering Workshops, 2005. 21st International Conference on. IEEE, 2005.

[4] Krumm, John. "Trajectory analysis for driving." Computing with Spatial Trajectories. Springer New York, 2011. 213-241.

[5] Li, Min, Zhiguang Qin, and Cong Wang. "Query-Privacy-Aware Location Cloaking for Mobile P2P System." International Journal Of Future Generation Communication And Networking , VOl 6, no. 4 August 2013

[6] Lu, Hua, Christian S. Jensen, and Man Lung Yiu. "Pad: Privacy-area aware, dummy-based location privacy in mobile services." Proceedings of the Seventh ACM

International Workshop on Data Engineering for Wireless and Mobile Access. ACM, 2008.

[7] Murugesan, Mummoorthy, and Chris Clifton. "Providing Privacy through Plausibly Deniable Search." SDM. 2009.

[8] Pingley, Aniket, et al. "Protection of query privacy for continuous location based services." INFOCOM, 2011 Proceedings IEEE. IEEE, 2011.

[9] Stenneth, Leon, P. S. Yu, and Ouri Wolfson. "Mobile systems location privacy:"MobiPriv" a robust k anonymous system." Wireless and Mobile Computing, Networking and Communications (WiMob), 2010 IEEE 6th International Conference on. IEEE, 2010.

[10] Xu N, Dan Zhu,Hongyan Liu, Jun He, Xiaoyong Du, Tao Liu. "Combinig Spatial Cloaking and Dummy Generation for Location privacy". Advanced Data Mining and Applications Lecture Notes in Computer Science Volume 7713, 2012, pp 701-712

[11] You, Tun-Hao, Wen-Chih Peng, and Wang-Chien Lee. "Protecting moving trajectories with dummies." Mobile Data Management, 2007 International Conference on. IEEE, 2007.

Ms. Priti Jagwani is pursuing PhD in Computer Science from School of IT, IIT Delhi. She has received her M.Tech degree in Computers from IIT Delhi in 2011. She is currently working as an Assistant Professor in Dept. of Computer Science, RLA (E) College, New Delhi, India. Her research interest is location privacy.

Ranjan Kumar has done his MCA from dept of Computer Science , Univ of Delhi. He is currently working as an assistant professor in Dept. of Computer Science, RLA (E) College, New Delhi, India. His research areas are sensor networks, location based services.