Towards a Scalable Model for Location Privacy

Nayot Poolsappasit
Computer Science Dept.
Colorado State University
Fort Collins, CO 80523
nayot@cs.colostate.edu

Indrakshi Ray
Computer Science Dept.
Colorado State University
Fort Collins, CO 80523
iray@cs.colostate.edu

ABSTRACT

With the growth of wireless and mobile technologies, we are witnessing an increase in location-based services (LBS). Although LBS provide enhanced functionalities, they open up new vulnerabilities that can be exploited to cause security and privacy breaches. Specifically, location data of individuals that are used by such services must be protected from security and privacy breaches. Such services will require new models for expressing privacy preferences for location data and mechanisms for enforcing them. We identify the factors on which location privacy depends and propose a scalable model for expressing privacy that can be used for LBS and other applications where the privacy of location information must be protected.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: [Security and Protection]

General Terms

Access control, privacy

Keywords

Location privacy

1. INTRODUCTION

Recent research on location technology has spawned numerous services, such as, FCC's Enhanced 911, AxisMobile's FriendZone, Verizon's Navigator, Sprint's Family Locator, RIM's Blackberry Service, or Intel's Thing Finder, that reveal location information with a high-degree of spatial precision. Such technology will not only provide enhanced functionality, but will also introduce additional security and privacy concerns. Specifically, location information of individuals subscribing to or using such services must be protected against security and privacy breaches. Models are needed that will allow individuals to express their privacy preferences and technologies are needed to enforce them.

Location-Based Services (LBS) can be classified into various categories based on the services they provide. Providing location privacy for all these different types of service may not make adequate use of resources. Some services, such as, point-of-interest service, do not require real-time information. Some services must verify legitimate use but may not need to know the exact identity or the precise location of the user. Examples include navigation service, local information service, and range queries. To protect the privacy of the user, researchers have proposed various approaches to blur the exact location of the user. Other services, such as friend finding, need both the identity of the user as well as his/her exact location. Privacy protection becomes critical for such applications.

The notion of privacy varies from one individual to another. One individual may be willing to disclose his location to his co-workers while he is on vacation, whereas another individual may not want to do so. The key question in location privacy is that who should have access to what location information and under what circumstances. Ideally, we need a model that will allow different users to express their location privacy preferences and mechanisms for enforcing them. Moreover, for reasons of implementation, the model should be scalable.

Developing a model that takes into account the personal privacy preferences of all potential individual users may not be very scalable. Towards this end, we have identified some factors that we feel are important for location privacy. These factors form the basis of our location privacy model. We propose three different models that use these factors for expressing privacy preferences. The models differ with respect to the computation requirements, and the granularity with which privacy preferences can be expressed. Finally, we also discuss implementation issues pertaining to our model.

The rest of the paper is organized as follows. Section 2 briefly enumerates the related work in this area. Section 3 describes the problem that we are trying to solve. Section 4 identifies the factors that are important to location privacy and proposes techniques for quantifying them. Section 5 discusses how location privacy can be enforced. Section 6 concludes the paper and mentions some future works.

2. RELATED WORK

IETF Geographic Location Privacy (GEO-PRIV) working group [3] addresses privacy and security issues pertaining to location information. They specify how location information

ACM SPRINGL '08, November 4,2008, Irvine,CA, USA.

can be transmitted in a secure manner and how the release of such information can be authorized. However, it does not address the issue of how to protect the privacy of the end user.

Gruteser and Grunwald [5] propose a spatio-temporal cloaking mechanism that allows the user's location to be indistinguishable from k people. In this scheme the trusted location intermediary gets the location update from all subscribed users and provides the service provider with the cloaked region that satisfies the user's k-anonymity.

The Clique-Cloak algorithm [4] takes a similar approach as [5] and builds a clique graph from a set of all subscribed users which is used to decide whether some users share the cloak spatial area. Due to the computation overhead of the clique graph, this approach does not scale very well. It becomes especially problematic for users that require a high value of k-anonymity.

New Casper was proposed in [7]. It employs a grid-based pyramid structure to index all user locations. The user can specify the value of k that dictates the level of anonymity required. They can also provide another parameter called *Amin* which gives the least granular location information that must be disclosed. New Casper uses a privacy-aware query processor to return a list of candidate query results to the anonymizing proxy, who has to locally refine the actual result from the candidate list. However, this approach still incurs high computation cost.

Kido et. al. propose the decentralized privacy protection with Dummy Location [6]. The main idea is that the user sends a set of false location called dummies along with the true location to LBS. The location server processes all requests and send all answers back to the subscriber. Then the client only pick one answer it desires from the candidate list. Clearly, the disadvantage of this approach is that the server wastes a lot of computation resource in processing false queries and the adversary may detect the true location from observing the request history.

Chi-Yin Chow et. al. propose a P2P spatial cloaking approach [1]. The proposed scheme exploits the anonymous peer-to-peer searching to construct the cloak region such that user cannot be distinguished from k other entities in the cloak area. Then the user selects an agent among the k entities who is responsible for forwarding the user's query to the service provider. This approach has some disadvantages. First, forming the group and selecting the agent may be challenging because not all mobile devices subscribe to the same service provider. Moreover, if a requester is malicious, the anonymous peer searching may breach location privacy of other devices.

In relating to the privacy policy, several social studies [2, 8, 9] were conducted with regards to the policy pertaining to the disclosure of private information. Palen et al. [8] found that the privacy management is a dynamic response to circumstances rather than a static enforcement of rules. They also emphasized that the social and institutional setting must be considered in developing the privacy-aware technology.

In response to [8], Consolvo [2], and Smith [9] conducted their individual research regarding the disclosure of location information. Their results all agree in that people making the decision of revealing their location regards to *who* is requesting, *why* they want to know the location, *when* and *where* the policy owner is, and *how* the policy owner feels about the requester at the time of request. The result of these social studies are used in our proposed privacy preference model.

Perhaps the closest work related to this paper is the Einar Snekkenes's location privacy model [10]. Snekkennes identify five components that play a major role in location privacy: requester, object, usage, time, and velocity and propose a lattice-based approach for location privacy. Since the complete lattice containing all information pertaining to the domains of the five components can be very large and is not very scalable, the author proposes to use a sparse lattice. This sparse lattice only covers circumstances that the policy owner anticipated. To handle the unexpected situations, the unforeseen scenario will be matched with the predefined circumstances that have at least one element in common. We believe that the lattice-based approach is too simplistic because the elements in the individual domains may not form a total order. For example, consider the requester domain. The element friend does not always dominate the element spouse – there are some circumstances where one may want the spouse to be the last person to know one's exact location. In addition, the paper does not address the what are the minimum requirements needed to build the lattice. Without this requirement, the initial information may be too sparse and have inadequate information to determine the preference on certain circumstances. This motivates us to develop a privacy preference model with primitive requirements so that the full policy can be generated from a minimum but adequate set of information.

3. PROBLEM STATEMENT

LBS are equipped to handle many different types of queries. Examples include "Find the gas stations within a radius of 2 miles", "Where is Smith?", or "Notify subscribers of the traffic on the street they are entering". Location-based queries have different entities associated with them. *Requester* is the entity who issues the location-based query. *Requested object* is the object or entity whose location is being queried. *Service provider* is responsible for providing services to customers. *Subscriber* is the customer who subscribes to services provided by the service provider. *Location provider* is the entity that computes the spatial information and is responsible for respecting the privacy of the location of the requested object. *Policy owner* is the entity who decides the location privacy of the requested object.

In this paper, we focus our attention to the queries where the requested object is a user or a device belonging to some user. Note that, for such queries, the location information must be disclosed in a controlled manner to protect the privacy and security of the individuals. The location provider should respect the privacy of the individual owner and provide information to the requester. It is the onus of the policy owner to specify what location information can be revealed to whom and under what circumstances. The factors that influence the willingness of an user to reveal his location information constitute the context of the query.

The response to a location query is the location information. Instead of giving the location in terms of physical coordinates, the system will respond with logical locations. Logical locations are the symbolic names associated with physical locations. Examples of logical locations are USA, Colorado, and Fort Collins. We assume that the system has an efficient mechanism for manipulating location data and

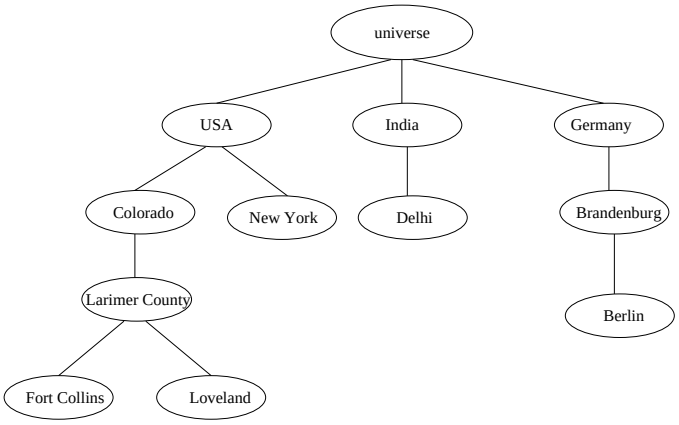

Figure 1: Example of a Location Hierarchy

translating physical locations to logical locations and vice-versa. The logical locations are organized in a hierarchical structure as shown in Figure 1. The nodes represent the different locations. The root of the hierarchy is the location *universe* which contains all other locations. If a node N_i appears higher up in the hierarchy and is connected to node N_j that appears lower in the hierarchy, we say that node N_i contains node N_j and is denoted by $N_j \subseteq N_i$. The hierarchical structure helps determine the location granularity. An user when specifying his privacy preference can choose the level of granularity at which he wishes to respond to the query.

The policy owner must provide his location privacy preferences. Location privacy depends on several factors (described in details in Section 4). These factors form the query context. The query context determines the location information that can be revealed to the user. Corresponding to the query context, we store information that specifies the details about location disclosure.

Definition 1 [**Location Privacy Preference**] The policy owner specifies the privacy preference as a set of tuples of the form $< c, loc_c >$ where c is the context of the query and loc_c is the location that is revealed in response to the query.

The response to a location query is said to be correct and privacy preserving if it satisfies several conditions. First, the actual most specific location of the object should be contained in the location that is returned in response to the query. Second, the location that is returned in response to the query should satisfy the location granularity and details that are specified in the privacy preference.

Definition 2 [**Privacy Preserving Location Response**] Let the context associated with the given query be c and loc_c is the location information associated with context c in the policy owner's privacy preference. Let loc_o be the most specific actual location of the requested object and loc_r be the response that is returned to the user. The location information returned to the user, loc_r, is said to be a *correct privacy preserving response* if $loc_o \subseteq loc_r$ and $loc_c \subseteq loc_r$.

4. FACTORS INFLUENCING LOCATION PRIVACY

In order to enforce location privacy, one needs to understand the factors that influence the willingness of an user to reveal his location information. First, the requester's identity or role plays an important part. An user may be willing to reveal his location information to his spouse but may not be willing to do so to strangers. Second, the usage information may also play a role for location privacy. An user may be willing to disclose his location information to volunteers during emergency operations, but may not do so otherwise. Third, the time when the information is requested also plays an important role. A person may reveal his location information to co-workers during his office hours, but may not do so during vacation. Fourth, location itself plays an important role in location privacy. A person may not be willing to reveal his location information when he is in the hospital undergoing some private treatment, but may reveal his location information when he is in the theater.

What makes location privacy a complex problem is the fact that the factors mentioned above are really not independent. Instead, location privacy depends on the combination of these factors. For example, a person may be willing to reveal his location information to his co-workers when he is in the office during the working day. Similarly, he may be unwilling to disclose his location information to his spouse when he is in the bar at midnight enjoying with his friends. The combination of these different factors form the context of the location-based query. The response provided by the user depends upon the context of the query.

Definition 3 The context formalizes the scenario under which a location query has been placed. The context of location query l is specified by the tuple $< I_l, U_l, T_l, L_l >$ where I_l represents the identity or role of the requester, U_l denotes the usage requirement of the requester, T_l specifies the time when the query is placed, L_l is the location of the requested object.

4.1 Representing the Factors

Since the context of each query has to be matched against the privacy preference of the user, we need a mechanism to represent each factor. For example, one may choose to represent the possible values for the factor identity as a set of strings. Similarly, the other factors can also be represented as sets of strings. The problem with this approach is that there has to be an exact match between the factors specified in the query context and those that are stored in the privacy preference profile. In the absence of an exact match, no response will be returned to the user.

The above problems can be removed to some extent if we quantify each factor in the context. The major advantage to such an approach is that it allows us to extrapolate context values for unknown circumstances. It also makes it easier to calculate the location preference for a given context. In the following, we describe how to assign numerical values to each factor.

4.1.1 Quantifying Requester's Role

Ideally, a person would like to reveal his location information based on the identity of the user. However, such a model will not scale well when there are a very large number of users. Thus, we propose using the role of the requester for determining location privacy. We identify certain important roles for location privacy. Examples include close relatives,

close friends, neighbors, co-workers, employers, adversaries, strangers, commercial agents, police, government workers etc. For each of these defined roles, we can adapt Bogardus social distance scale to measure relationship closeness. We assign a value between 0 and 1 for each such role. The value is near to 1 for close relationships and approximates 0 for remote relationships. Certain roles which may not represent close relationships may also be assigned a high value due to the nature of the role. Examples include social worker or law enforcement officer. The reason is that these roles must have access to location information.

4.1.2 Quantifying Usage

The requester must also specify how he is going to use the location information. All potential forms of usage can be organized in the form of a hierarchy. The nodes higher up in the hierarchy signify more general usage than those found lower in the hierarchy. The leaf nodes are assigned values in the range 0 to 1. 0 signifies that the usage is not very important and so information must not be disclosed. 1 signifies legitimate use and must be disclosed. The values for the intermediate node is calculated by taking the minimum value from its children. The process is repeated for the entire hierarchy.

4.1.3 Quantifying Time

The temporal attribute is also an important factor in location privacy. Time can also be represented in the form of a hierarchy. The root of the hierarchy is denoted as *always*. At the next level, we have working hours and non-working hours. Since location privacy is relatively less important during working hours, a value close to 1 is assigned. For non-working hours location privacy may be extremely important and a value close to 0 may be assigned.

4.1.4 Quantifying Location

The propensity to disclose location information may be dependent on location itself. We can organize location in the form of a hierarchy and associate values with it. The values as before range from 0 to 1. Nodes higher up in the hierarchy are assigned a greater value than nodes lower down. This is because a user may be more willing to disclose less granular location information. However, nodes within a level may be assigned different values depending on the sensitivity. For example, the nodes hospital and park have different values associated with them because they differ in sensitivity.

5. ENFORCING LOCATION PRIVACY

A naive approach that works with all kinds of representation is to build a list of all the possible contexts and associating a level of location disclosure with it. The context of the query posed by the user is matched with the set of contexts stored and the corresponding location information is returned to the user. The advantage of such an approach is that it is simple and gives the accurate location disclosure preference in the case of an exact match. Such a naive approach has several problems. First, we need to identify all possible contexts and associate location preferences for them. Second, the number of entries may be very large and it may not be efficient to search through them. Third, if the context of the query does not match any of the entries, the requester will not receive any information.

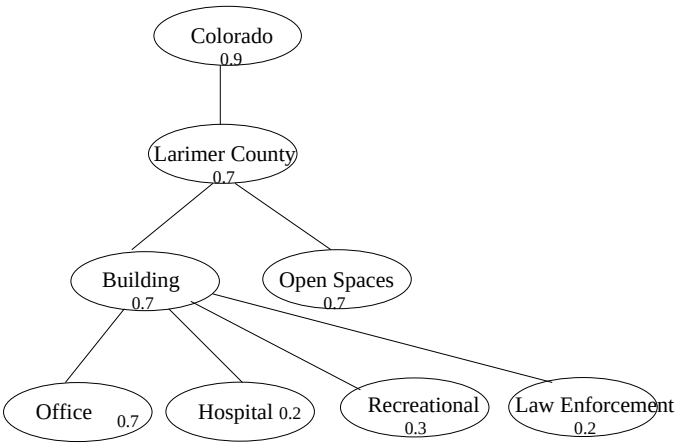

Figure 4: Location Hierarchy

Once numerical values have been assigned to the different factors, we can have different techniques for calculating privacy preferences. One approach is to assign weights to each factor based on the preference. Let w_i, w_u, w_t and w_l be the weights assigned to the factors, namely, identity, usage, time, and location respectively. Note that, $w_i + w_u + w_t + w_l = 1$ and $0 \leq w_i, w_u, w_t, w_l \leq 1$. The value of each factor is computed from the query context. Let v_i, v_u, v_t, and v_l be the values obtained for each factor specified in the query context. The level of privacy preference is computed as follows. $L_p = w_i * v_i + w_u * v_u + w_t * v_t + w_l * v_l$. The granularity at which location information can be disclosed is a function of the privacy preference. Higher values of L_p correspond to specifying location information at finer granularity. The level of privacy, L_p, is an input to the blurring function. The blurring function will be used to return the location information at the appropriate granularity level. The advantage of this approach is that it is easy computationally. The problem with the above approach is that it requires the user to assign preferences to each of the factors. It is not always possible to form a total order among all the factors. For instance, for some requesters, the location of the user may have a higher importance than time of the day. For other requesters, it may be the opposite.

The next approach that we propose is a little different. Among all the factors that influence location privacy, role of the requester is perhaps the most important. We propose a scheme in which the policy owner considers three types of combinations: requester and usage, requester and time, and requester and location. For each of these combinations, he specifies his preference to disclose location information. In other words, we define three functions: $T_u : I \times U \rightarrow P$, $T_t : I \times T \rightarrow P$, and $T_l : I \times L \rightarrow P$ where $P \in [0, 1]$. The preference value 0 indicates the policy owner's unwillingness to disclose location information, and 1 indicates complete willingness to disclose location information. This allows each user to assign preferences to the combination of the requester and usage, requester and time, and requester and location. The importance of each combination is denoted by the weight factor. Let w_u, w_t and w_l be the three weights associated with the usage, time, and location factors corresponding to a given requester i. Here $0 \leq w_u, w_t, w_l \leq 1$ and $w_u + w_t + w_l = 1$. Let pu_{ij} be the preference associated

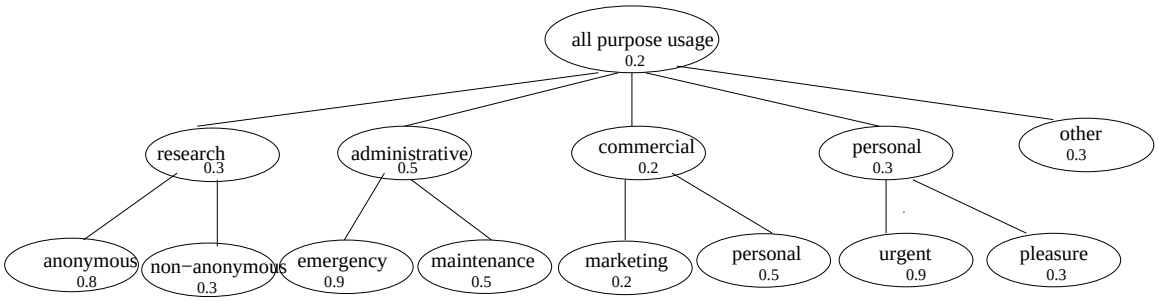

Figure 2: Usage Hierarchy

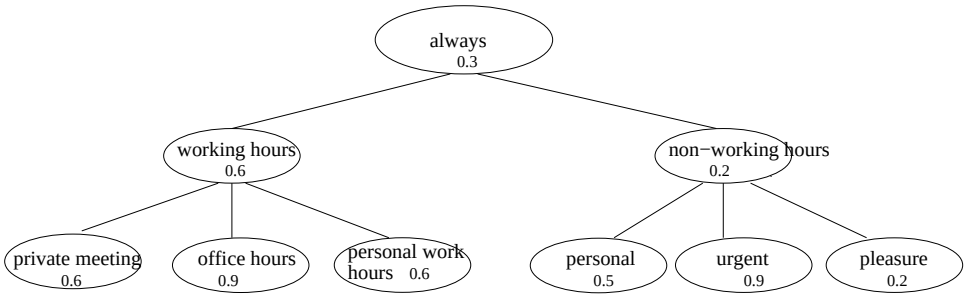

Figure 3: Temporal Hierarchy

with requester i and usage j, pt_{ik} be the preference associated with requester i and time k, and pl_{im} be the preference associated with requester i and location m. The level of location privacy L_p is given by $L_p = w_u{*}pu_{ij}+w_t{*}pt_{ik}+w_l{*}pl_{im}$. Here again, we use the L_p as an input to the blurring function to return the location information at the correct granularity.

6. CONCLUSION

Technological advancements in mobile computing has spawned a growth in location-based services. Such services use the location information of the subscriber to provide better functionalities. Improper usage of location information may compromise the security and privacy of an individual. Moreover, a user must be allowed to control who has access to his location information and under what circumstances. Towards this end, we investigate the factors influencing location privacy, suggest techniques for quantifying them, and propose different approaches for expressing the user's privacy preference with respect to the disclosure of location information. The approaches differ with respect to the storage requirements, and the granularity of privacy preference. A lot of work remains to be done. First, we need to look at the implementation issues pertaining to the model. Specifically, how is location information stored and managed in our model. Next, we need to validate our model using real-world applications and data.

7. ACKNOWLEDGEMENTS

The work was supported in part by AFOSR under contract number FA9550-07-1-0042.

8. REFERENCES

[1] C.Y. Chow, M.F. Mokbel, and X. Liu. A Peer-to-Peer Spatial Cloaking Algorithm for Anonymous Location-based Service. *Proceedings of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems*, pages 171–178, 2006.

[2] S. Consolvo, I. E. Smith, T. Matthews, A. LaMarca, J. Tabert, and P. Powledge. Location disclosure to social relations: why, when, & what people want to share. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 81–90, New York, NY, USA, April 2005. ACM Press.

[3] J. Cuellar, J. Morris, and D. Mulligan. RFC 4079: Geopriv requirements. *Internet Engineering Task Force (IETF) Internet Draft. http://www. ietf. org/ids. by. wg/geopriv. html*, 2003.

[4] B. Gedik and L. Liu. Location Privacy in Mobile Systems: A Personalized Anonymization Model . *Proceedings of the 25th International Conference on Distributed Computing Systems*, pages 620–629, 2005.

[5] M. Gruteser and D. Grunwald. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*, pages 31–42, 2003.

[6] H. Kido, Y. Yanagisawa, and T. Satoh. An Anonymous Communication Technique using Dummies for Location-based Services. *Proceedings of IEEE International Conference on Pervasive Services*, pages 88–97, 2005.

[7] M.F. Mokbel, C.Y. Chow, and W.G. Aref. The New Casper: Query Processing for Location Services without Compromising Privacy. *Proceedings of the 32nd International Conference on Very Large Data Bases*, pages 763–774, September 2006.

[8] L. Palen and P. Dourish. Unpacking" privacy" for a

networked world. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 129–136, New York, NY, USA, 2003. ACM Press.

[9] I. Smith, S. Consolvo, J. Hightower, J. Hughes, G. Iachello, A. LaMarca, J. Scott, T. Sohn, and G. Abowd. Social Disclosure of Place: From Location Technology to Communication Practice. *Proceedings of the 3rd International Pervasive Computing Conference*, pages 134–151, 2005.

[10] E. Snekkenes. Concepts for personal location privacy policies. In *Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 48–57, New York, NY, USA, 2001. ACM Press.