International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 14 Issue 1 –APRIL 2015.

Exploiting Service Similarity for Privacy in Location-Based Search Queries

M.SIVANESAN, *M.Tech IT, Sona College of Technology, Salem*
msivanesanmtech@gmail.com
J.ALDO STALIN, *M.E, Assistant Professor, Sona College of Technology, Salem*

ABSTRACT
Location-based applications utilize the positioning capabilities of a mobile device to determine the current location of a user, and customize query results to include neighboring points of interests. However, location knowledge is often perceived as personal information. One of the immediate issues hindering the wide acceptance of location-based applications is the lack of appropriate methodologies that offer fine grain privacy controls to a user without vastly affecting the usability of the service. While a number of privacy-preserving models and algorithms have taken shape in the past few years, there is an almost universal need to specify one's privacy requirement without understanding its implications on the service quality. In this paper, we propose a user-centric location-based service architecture where a user can observe the impact of location inaccuracy on the service accuracy before deciding the geo-coordinates to use in a query. We construct a local search application based on this architecture and demonstrate how meaningful information can be exchanged between the user and the service provider to allow the inference of contours depicting the change in query results across a geographic area. Results indicate the possibility of large default privacy regions (areas of no change in resultset) in such applications.

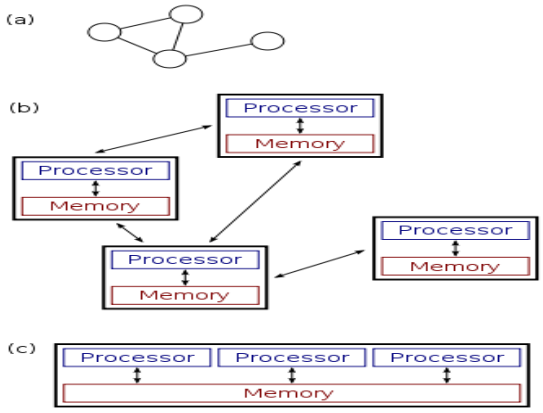
INTRODUCTION

Distributed computing is a field of computer science that studies distributed systems. A distributed system is a software system in which components located on networked computers communicate and coordinate their actions by passing messages. The components interact with each other in order to achieve a common goal. There are many alternatives for the message passing mechanism, including RPC-like connectors and message queues. Three significant characteristics of distributed systems are: concurrency of components, lack of a global clock, and independent failure of components. An important goal and challenge of distributed systems is location transparency. Examples of distributed systems vary from SOA-based systems to massively multiplayer online games to peer-to-peer applications. A computer program that runs in a distributed system is called a distributed program, and distributed programming is the process of writing such programs. Distributed

computing also refers to the use of distributed systems to solve computational problems. In distributed computing, a problem is divided into many tasks, each of which is solved by one or more computers, which communicate with each other by message passing. The word *distributed* in terms such as "distributed system", "distributed programming", and "distributed algorithm" originally referred to computer networks where individual computers were physically distributed within some geographical area. The terms are nowadays used in a much wider sense, even referring to autonomous processes that run on the same physical computer and interact with each other by message passing. While there is no single definition of a distributed system, the following defining properties are commonly are several autonomous computational entities, each of which has its own local memory, the entities communicate with each other by message passing. In this article, the computational entities are called *computers* or *nodes*. A distributed system may have a common goal, such as solving a large computational problem.[1] Alternatively, each computer may have its own user with individual needs, and the purpose of the distributed system is to coordinate the use of shared resources or provide communication services to the users.

Other typical properties of distributed systems include the following: The system has to tolerate failures in individual computers. The structure of the system (network topology, network latency, number of computers) is not known in advance, the system may consist of different kinds of computers and network links, and the system may change during the execution of a distributed program. Each computer has only a limited, incomplete view of the system. Each computer may know only one part of the input. Distributed systems are groups of networked computers, which have the same goal for their work. The terms "concurrent computing", "parallel computing", and "distributed computing" have a lot of overlap, and no clear distinction exists between them. The same system may be characterised both as "parallel" and "distributed"; the processors in a

typical distributed system run concurrently in parallel. Parallel computing may be seen as a particular tightly coupled form of distributed computing, and distributed computing may be seen as a loosely coupled form of parallel computing. Nevertheless, it is possible to roughly classify concurrent systems as "parallel" or "distributed" using the following criteria: In parallel computing, all processors may have access to a shared memory to exchange information between processors. In distributed computing, each processor has its own private memory (distributed memory). Information is exchanged by passing messages between the processors.



The figure on the right illustrates the difference between distributed and parallel systems. Figure (a) is a schematic view of a typical distributed system; as usual, the system is represented as a network topology in which each node is a computer and each line connecting the nodes is a communication link. Figure (b) shows the same distributed system in more detail: each computer has its own local memory, and information can be exchanged only by passing messages from one node to another by using the available communication links. Figure (c) shows a parallel system in which each processor has a direct access to a shared memory. The situation is further complicated by the traditional uses of the terms parallel and distributed *algorithm* that do not quite match the above definitions of parallel and distributed *systems*; see the section Theoretical foundations below for more detailed discussion. Nevertheless, as a rule of thumb, high-performance parallel computation in a shared-memory multiprocessor uses parallel algorithms while the coordination of a large-scale distributed system uses distributed algorithms.

SURVEY ANALYSIS

On the Anonymity of Home/Work Location Pairs for an applications benefit from user location data, but location data raises privacy concerns. Anonymization can protect privacy, but identities can sometimes be inferred from supposedly anonymous data. This paper studies a new attack on the anonymity of location data. We show that if the approximate locations of an individual's home and workplace can both be deduced from a location trace, then the median size of the individual's anonymity set in the U.S. working population is 1, 21 and 34,980, for locations known at the granularity of a census block, census track and county respectively. The location data of people who live and work in different regions can be re-identified even more easily. Our results show that the threat of re-identification for location data is much greater when the individual's home and work locations can both be deduced from the data. To preserve anonymity, we offer guidance for obfuscating location traces before they are disclosed.

Anonymization of Location Data Does Not Work: A Large-Scale Measurement Study, a very large-scale data set of more than 30 billion call records made by 25 million cell phone users across all 50 states of the US and attempt to determine to what extent anonymized location data can reveal private user information. Our approach is to infer, from the call records, the "top N" locations for each user and correlate this information with publicly-available side information such as census data. For example, the measured "top 2" locations likely correspond to home and work locations, the "top 3" to home, work, and shopping/school/commute path locations. We consider the cases where those "top N" locations are measured with different levels of granularity, ranging from a cell sector to whole cell, zip code, city, county and state. We then compute the anonymity set, namely the number of users uniquely identified by a given set of "top N" locations at different granularity levels. We find that the "top 1" location does not typically yield small anonymity sets. However, the top 2 and top 3 locations do, certainly at the sector or cell-level granularity. We consider a variety of different factors that might impact the size of the anonymity set, for example the distance between the "top N" locations or the geographic environment (rural vs urban). We also examine to what extent specific side information, in particular the size of the user's social network, decrease the anonymity set and therefore increase risks to privacy. Our study shows that sharing anonymized location data will likely lead to privacy risks and that, at a minimum, the data needs to be coarse in either the time domain (meaning the data is collected over short periods of time, in which case inferring the top N locations reliably is difficult) or the space domain (meaning the data granularity is

strictly higher than the cell level). In both cases, the utility of the anonymized location data will be decreased, potentially by a significant amount.

A Formal Model of Obfuscation and Negotiation for Location Privacy the practice of deliberately degrading the quality of information in some way, so as to protect the privacy of the individual to whom that information refers. In this paper, we argue that obfuscation is an important technique for protecting an individual's location privacy within a pervasive computing environment. The paper sets out a formal framework within which obfuscated location-based services are defined. This framework provides a computationally efficient mechanism for balancing an individual's need for high-quality information services against that individual's need for location privacy. Negotiation is used to ensure that a location-based service provider receives only the information it needs to know in order to provide a service of satisfactory quality. The results of this work have implications for numerous applications of mobile and location-aware systems, as they provide a new theoretical foundation for addressing the privacy concerns that are acknowledged to be retarding the widespread acceptance and use of location-based services.

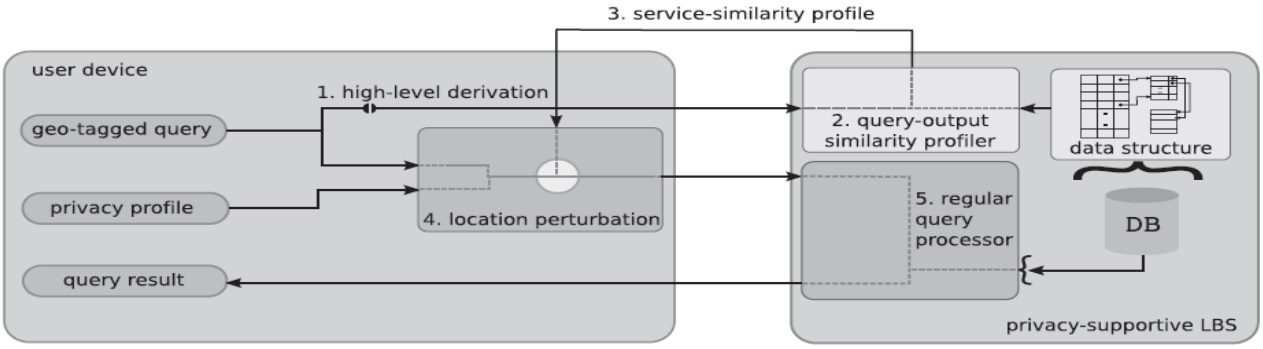
An Anonymous Communication Technique Using Dummies for Location-Based Services for highly accurate positioning devices enable us to provide various types of location-based services. On the other hand, because such position data include deeply personal information, the protection of location privacy is one of the most significant problems in location-based services. In this paper, we propose an anonymous communication technique to protect the location privacy of the users of location-based services. In our proposed technique, such users generate several false position data (dummies) to send to service providers with the true position data of users. Because service providers cannot distinguish the true position data, user location privacy is protected. We also describe a cost reduction technique for communication between a client and a server. Moreover, we conducted performance study experiments on our proposed technique using practical position data. As a result of the experiments, we observed that our proposed technique protects the location privacy of people and can sufficiently reduce communication costs so that our communication techniques can be applied in practical location-based services.

Preserving User Location Privacy in Mobile Data Management Infrastructures for Location-based services, such as finding the nearest gas station, require users to supply their location information. However, a user's location can be tracked without her consent or knowledge. Lowering the spatial and temporal resolution of location data sent to the server has been proposed as a solution. Although this technique is effective in protecting privacy, it may be overkill and the quality of desired services can be severely affected. In this paper, we suggest a framework where uncertainty can be controlled to provide high quality and privacy-preserving services, and investigate how such a framework can be realized in the GPS and cellular network systems. Based on this framework, we suggest a data model to augment uncertainty to location data, and propose imprecise queries that hide the location of the query issuer and yields probabilistic results. We investigate the evaluation and quality aspects for a range query. We also provide novel methods to protect our solutions against trajectory-tracing. Experiments are conducted to examine the effectiveness of our approaches.

METHODOLOGY USED

In existing system for location obfuscation has been extensively investigated in the context of privacy. Obfuscation has been earlier achieved either through the use of dummy queries or cloaking regions. In the dummy query method, a user hides her actual query (with the true location) among a set of additional queries with incorrect locations. Data model to augment uncertainty to location data using circular regions around all objects. They use imprecise queries that hide the location of the query issuer and yield probabilistic results. Gedik and Liu develop a location privacy architecture where each user can specify maximum temporal and spatial tolerances for the cloaking regions. Drawing inspiration from the concept of k-anonymity in database privacy, Gedik and Liu enforce a location k-anonymity requirement while creating the cloaking regions. This requirement ensures that the user will not be uniquely located inside the region in a given period of time. Ghinita et al. propose a decentralized architecture to construct an anonymous spatial region, and eliminate the need for the centralized anonymizer. In their approach, mobile nodes utilize a distributed protocol to self-organize into a fault-tolerant overlay network, from which a k-anonymous cloaking set of users can be determined. The disadvantages of existing system are Parameter specification remains the biggest hindrance to real-world application of these techniques. Even when a user has advanced knowledge to comprehend the implications of a parameter setting on location privacy, the impact on service is unknown in these approaches. The lack of appropriate methodologies

that offer fine grain privacy controls to a user without vastly affecting the usability of the service.



Architecture Diagram for Location Based Search Queries

The Proposed System having a novel architecture to help identify privacy and utility tradeoffs in an LB The architecture has a user-centric design that delays the sharing of a location coordinate until the user has evaluated the impact of its accuracy on the service quality. Precise geo-locations are necessary for result set accuracy when the queried objects exist as a dense cluster in the search area.
The Advantages Of Proposed System using approach can precisely reveal the area boundaries within which the result set is fully preserved (a default privacy level). A high degree of precision in estimating the area boundaries when user requirements on result set accuracy are relaxed (i.e., location sensitivity is hardened) and differencing algorithms can be used to reduce the communication overhead.

CONCLUSION

Based on the observations from the empirical study, we make the following conclusions on the efficacy of a privacy supportive local search application. Precise geo locations are necessary for result set accuracy when the queried objects exist as a dense cluster in the search area. It seems unlikely that both location privacy and result exactness can be maintained in this case. A privacy supportive application would allow the user to aggressively tradeoff the service similarity requirement to determine a sufficiently large area for location perturbation. Given the high density of objects, resulting objects can still be expected to be in the near vicinity. When object density is not dense, location accuracy has a minor role to play in retrieving relevant results. A privacy supportive application would help identify the large default-privacy regions resulting in such situations. Next generation telecommunication systems could very well make it possible to quickly (and cost-effectively) transfer all information required to infer the service contour exactly. Until then, approximate inferencing

algorithms can be used to reduce the communication overhead.

REFERENCES

[1] J. Sythoff and J. Morrison, Location-Based Services: Market Forecast, 2011-2015, Pyramid Research, 2011.

[2] P. Golle and K. Partridge, "On the Anonymity of Home/Work Location Pairs," Proc. Seventh Int'l Conf. Pervasive Computing, pp. 390-397, 2009.

[3] H. Zang and J. Bolot, "Anonymization of Location Data Does Not Work: A Large-Scale Measurement Study," Proc. 17th Ann. Int'l Conf. Mobile Computing and Networking, pp. 145-156, 2011.

[4] M. Duckham and L. Kulik, "A Formal Model of Obfuscation and Negotiation for Location Privacy," Proc. Third Int'l Conf. Pervasive Computing, pp. 152-170, 2005.

[5] H. Kido, Y. Yanagisawa, and T. Satoh, "An Anonymous Communication Technique Using Dummies for Location-Based Services," Proc. IEEE Int'l Conf. Pervasive Services, pp. 88-97, 2005.

[6] R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar, "Preserving User Location Privacy in Mobile Data Management Infrastructures," Proc. Sixth Workshop Privacy Enhancing Technologies, pp. 393-412, 2006.

[7] M.L. Yiu, C.S. Jensen, X. Huang, and H. Lu, "SpaceTwist: Managing the Trade-Offs among Location Privacy, Query Performance, and Query Accuracy in Mobile Services," Proc. 24th Int'l Conf. Data Eng., pp. 366-375, 2008.

[8] M. Gruteser and D. Grunwald, "Anonymous Usage of Location- Based Services through Spatial and Temporal Cloaking," Proc. First Int'l Conf. Mobile Systems, Applications, and Services, pp. 31-42, 2003.

[9] B. Gedik and L. Liu, "Protecting Location Privacy with Personalized k-Anonymity: Architecture and Algorithms," IEEE Trans. Mobile Computing, vol. 7, no. 1, pp. 1-18, Jan. 2008.

[10] P. Samarati, "Protecting Respondents' Identities in Microdata Release," IEEE Trans. Knowledge and Data Eng., vol. 13, no. 6, pp. 1010-1027, Nov. 2001.

[11] G. Ghinita, P. Kalnis, and S. Skiadopoulos, "PRIVE: Anonymous Location-Based Queries in Distributed Mobile Systems," Proc. 16th Int'l Conf. World Wide Web, pp. 371-380, 2007.

[12] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias, "Preventing Location-Based Identity Inference in Anonymous Spatial Queries," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 12, pp. 1719-1733, Dec. 2007.

[13] G. Ghinita, K. Zhao, D. Papadias, and P. Kalnis, "A Reciprocal Framework for Spatial k-Anonymity," J. Information Systems, vol. 35, no. 3, pp. 299-314, 2010.

[14] P.K. Agarwal, M. de Berg, J. Matousek, and O. Schwarzkopf, "Constructing Levels in Arrangements and Higher Order Voronoi Diagrams," Proc. 10th Ann. Symp. Computational Geometry, pp. 67- 75, 1994.

[15] F. Aurenhammer and O. Schwarzkopf, "A Simple On-line Randomized Incremental Algorithm for Computing Higher Order Voronoi Diagrams," Proc. Seventh Ann. Symp. Computational Geometry, pp. 142-151, 1991.

[16] D.-T. Lee, "On k-Nearest Neighbor Voronoi Diagrams in the Plane," IEEE Trans. Computers, vol. C-31, no. 6, pp. 478-487, June 1982.

[17] K.V. Mardia, "Some Properties of Classical Multidimensional Scaling," Comm. Statistics - Theory and Methods, vol. A, no. 7, pp. 1233-1241, 1978.

[18] A. Beygelzimer, S. Kakade, and J. Langford, "Cover Trees for Nearest Neighbor," Proc. 23rd Int'l Conf. Machine Learning, pp. 97- 104, 2006.