A Framework For Enhancing Privacy In Location Based Services Using K-Anonymity Model

Jane Mugi, Michael Kimwele, George Okeyo.

Abstract: This paper presents a framework for enhancing privacy in Location Based Services using K-anonymity model. Users of location based services have to reveal their location information in order to use these services; however this has threatened the user privacy. K-anonymity approach has been studied extensively in various forms. However, it is only effective when the user location is fixed. When a user moves and continuously sends their location information, the location service provider can approximate user trajectory which poses a threat to the trajectory privacy of the user. This framework will ensure that user privacy is enhanced for both snapshot and continuous queries. The efficiency and effectiveness of the proposed framework was evaluated, the results indicate that the proposed framework has high success rate and good run time performance.

Index Terms: Anonynizer, Anonymized Spatial Region, K-Anonymity, Location Based Services, Point of Interest.

1.0 INTRODUCTION

Due to the rapid advances in positioning technologies such as GPS, GSM, RFID and Wi-Fi (802.11b/g/n), mobile devices are often equipped with geo-located and wireless communication capacities. These recent development of ubiquitous devices have led to the development of a new class of services known as Location Based Services, that are tailored to the current location of the individual querying the service. LBS can be defined as a service that takes as input the current location of a user (generally acquired through a mobile device carried by this user) and tailors its output depending on the acquired location data. LBS can access, combine, and transform contextual information and more specifically location information, in order to personalize the service provided to the user. For example, LBS can be used for resource discovery, path finding, real time social applications or location-based gaming. When people use LBS to support them in their daily tasks, their position is usually acquired automatically through mobile equipments they carry with them, thus these systems continuously monitor and reveal information about the location of their users as the position of these mobile systems is essentially the same as the users of such systems[2]. Users with location-aware mobile devices can issue location-based snapshot or continuous queries to a database server at anytime and anywhere. Examples of snapshot queries include "Where my nearest gas station is" and "what are the restaurants within one mile of my location", Users with location-aware mobile devices can issue location-based snapshot or continuous queries to a database server at anytime and anywhere. Examples of snapshot queries include "Where my nearest gas station is" and "what are the restaurants within one mile of my location", while examples of continuous queries include "continuously report my nearest police car" and continuously report the taxis within one mile of my car. Although location-based services promise safety and convenience, they threaten the security and privacy of their customers. With untrustworthy servers, an adversary may access sensitive information about an individual's based on their issued location-based queries. E.g. an adversary may check a user's habit and interest by knowing the places she seeks [3]. Due to the nature of spatial queries, LBS needs the user position in order to process her requests. LBS makes spatial data available to the users through one or more location servers (LS) that index and answer user queries on them. Examples of spatial queries could be "where is the closest hospital to my current location?" or which pharmacies are open within a 1km radius? In order for the LS to be able to answer such questions, it needs to know the position of the querying user. There exist many algorithms for efficient spatial query processing, but the main challenge in the LBS industry is of a different nature. In particular, users are reluctant to use LBSs, since revealing their position may link to their identity. Even though a user may create a fake ID to access the service, her location alone may disclose her actual identity. Linking a position to an individual is possible by various means such as publicly available information such as telephone directories. User privacy may be threatened because of the sensitive nature of accessed data e.g. inquiring for pharmacies that offer medicines for diseases associated with a social stigma, or asking for nearby addiction recovery groups. Another source of threats comes from less sensitive data e.g. gas station, shops, restaurants, that may reveal the user's interest and shopping needs, resulting in a flood of unsolicited advertisements through e-coupons and personal messages [4].

- *Jane Mugi, School of Computing and Information Technology, Jomo Kenyatta University of Agriculture and Technology, P. O. BOX 62000-00200 Nairobi, Kenya,Email: jmugi@jkuat.ac.ke*
- *Michael Kimwele, School of Computing and Information Technology, Jomo Kenyatta University of Agriculture and Technology, P. O. BOX 62000-00200 Nairobi, Kenya, Email: kimwele@icsit.jkuat.ac.ke*
- *George Okeyo,School of Computing and Information Technology, Jomo Kenyatta University of Agriculture and Technology, P. O. BOX 62000-00200 Nairobi, Kenya, Email: gokeyo@jkuat.ac.ke*

2.0 LBS Techniques and Models

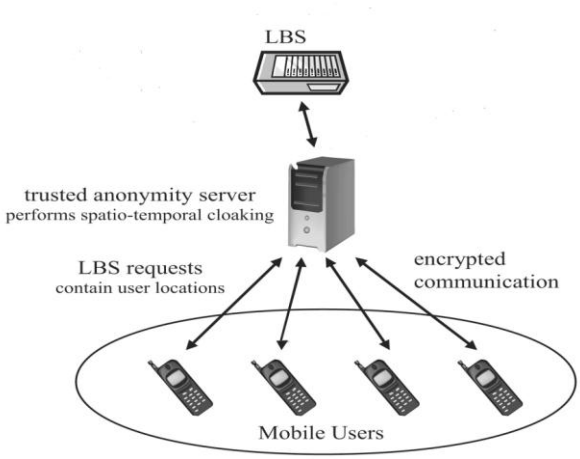
Mix zones approach was proposed by [1] to define areas called mix zones, where all user positions must be hidden such that the user position is not known within these zones This is achieved by not sending any position updates within a zone. It has the advantage of location and sampling accuracy but operation lack in multiple resp onder. Dummy-Q approach was proposed by [25] in which a user sends the actual location with several fake locations ("dummies") to a service provider. The challenge of dummy-Q is that 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. E.g. while users on a coastal location may often query for beaches, the same service attribute value may be quite rare around a desert rea. If such adherence to the trend of queries is shunned then the adversary may be able to exclude certain service attributes according to common sense and thereby identify the real query. The Space twist framework aims to offer location privacy for K nearest neighbor (kNN) queries at low communication cost without requiring a trusted anonymizer.The client specifies a fake user location called an anchor, which utilizes incremental NN query processing at the server. The server returns data points to the user incrementally in ascending order of their distances from the anchor [26]. Space twist rectifies the shortcomings of k nearest neighbor queries. This approach is flexible, needs no trusted middleware and requires only well-known incremental NN query processing on the server. Space twist ensures that no duplicates are retrieved. It offers a server side ring ranking technique that reduces the communication cost of exact queries. A delayed termination technique that reduces the communication cost of exact queries. Application to spatial networks. However Space Twist may fail since it cannot guarantee K-anonymity. The New Casper is a framework in which mobile and stationary users can entertain location-based services without revealing their location information. Casper consists of two main components the location anonymizer and the privacy-aware query processor. The location anonymizer hides the users exact location information into cloaked spatial regions based on user-specified privacy requirements. The privacy-aware query processor is integrated inside the location-based database server in order to deal with the cloaked spatial areas rather than the exact location information. Casper maintains an anonymizer and a privacy aware query processor. For a successful anonymization and cloaking, a pyramid structure is maintained. The pyramid structure is dynamically maintained to keep track of the current number of mobile users within each cell. Thus updating and cloaking is very expensive due to maintaining the pyramid structure.

3.0 Proposed Framework

The proposed work is focused on the K-anonymity model which uses trusted third party architecture since these systems are now being deployed to the public [27].The trusted third party model utilizes the concept of a middleware between the mobile user and the LBS. We sometimes refer to the middleware as an anonymization server or *AS*. Mobile requests are first sent to the middleware, the incoming request is then cloaked with other requests by the anonymization server before submission to the LBS. The proposed work is focused on the trusted third party architecture since these systems are now being deployed to the public. Mobile clients communicate with third-party LBS providers through the anonymity server. The anonymity server is a secure gateway to the LBS providers for the mobile clients. Each message sent to an LBS provider contains the location information of the mobile client and a time stamp, in addition to service-specific information. Upon receiving a message from a mobile client, the anonymity server hides the location information through spatio-temporal cloaking, and then forwards the anonymized message to the LBS provider.

Figure 3.0 *Trusted Third Party Architecture.*



4.0 Experimental Analysis and Results.

To evaluate the efficiency and effectiveness of our framework the following evaluation criteria was used.

4.1 Success Rate

Success rate is one of the most important evaluation criteria. The main goal of any anonymization server is to maximize the number of messages that can be successfully anonymized with the personalize quality of service and privacy requirement desired. The *success rate* is measured as the ratio of the number of successful anonymized request, by the total number of incoming mobile request. A success rate of 100% implied that all the requests that were sent by the mobile clients were safely anonymized. In the previous work on new Casper framework the request is dropped because the privacy requirement cannot be satisfied [3]. Let *N* be the total number of mobile requests sent by mobile clients to the anonymization server. The number of mobile request that can be anonymized successfully by the anonymization server is Mt as shown by (1).

$$N/M_t * 100 = 100\% \quad (1)$$

This implies that all the queries sent to the system will be anonymized successfully and none will be dropped. In our evaluations we refer to this property as the *success rate* of the proposed framework.

4.2 Perfomance Measure

An algorithm with a lower cloaking time does better, because the cloaking time is a measure of the temporal complexity. Efficient cloaking implies that the algorithm spends less time processing the incoming mobile requests from the mobile clients We define a function startTime (cloaking Algorithm), that returns the time the cloaking algorithm start the anonymization process. Also, endTime (cloaking Algorithm), which returns the time the cloaking algorithm completes the anonymization process as shown by (2).

Cloakingtime=endTime(CloakM(ms))-startTime(CloakM(ms)

4.3 Experimental Setup.

The experiment was conducted on a windows machine running Pentium(R) duo-core cpu 2.10 GHz processor with

4GB of ram. Jmeter software was used to simulate a group of 100 users to send requests to the anonymizer server. Simulation was done starting with 10 mobile requests adding 10 mobile request after every request and observing the performance until 100 mobile request were sent. The statistics were returned that showed the performance through graphs and tables as shown below.

4.4 Experimental results for Success Rate.

The experimental results for success rate are shown through the Jmeter summary report that was generated when 80 mobile request were simulated to the server. In most of the request that was made the error % was zero, because most of the requests ran successfully as shown in the report below. Only a few request showed a slightly lower error rate. The success rate of most of the mobile request sent to the server was 100%.

NO OF MOBILE REQUEST	AVERAGE RESPONSE TIME	MAX TIME	MIN TIME	ERROR %
10	3498MS	6749MS	956MS	0.00%
20	1246MS	3580MS	563MS	0.00%
30	6136MS	11250MS	564MS	0.00%
40	4621MS	52773MS	563 MS	0.00%
50	8877MS	113129MS	564MS	0.00%
60	12704MS	463022MS	564MS	0.00%
70	6313MS	140468MS	581MS	1.43%
80	8424MS	122276MS	579MS	0.00%
90	6684MS	119033MS	572MS	1.11%
100	6232	57962MS	568MS	2.00%

Time in MS

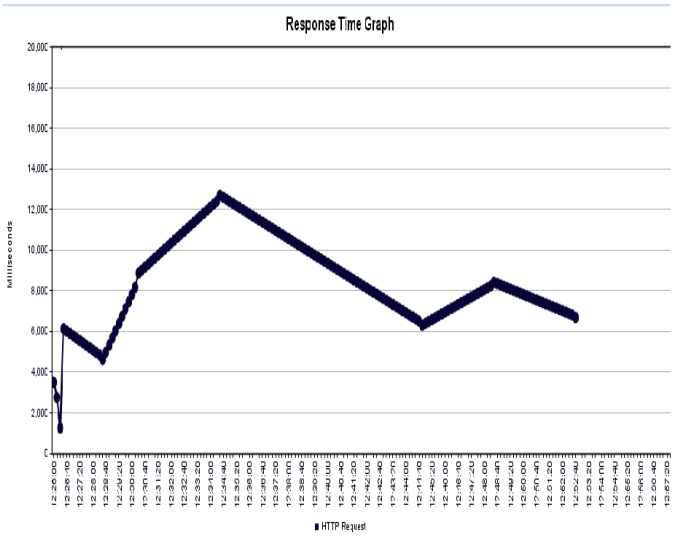
Table4.0 *Generated Summary report of 100 mobile requests.*

Average: Average is the average response time for that particular http request. This response time is in milliseconds.
Min: Min denotes to the minimum response time taken by the mobile request.
Max: Max denotes to the maximum response time taken by the mobile request.
Error %: This denotes the error percentage in samples during run.Most of the mobile request generated 0.00% error, this implies that most of the request ran successfully.

4.4.2 Experimental results for Performance Measure.

The experimental results for success rate are shown through the response time graph that was generated after simulating request to the server. The run time performance of algorithms is measured as the cloaking time. An algorithm with a lower cloaking time does better as shown by (2).
Cloakingtime=endTime(CloakM(ms))–startTime(CloakM(ms).
The graph below shows the response time when 100 mobile requests were sent to the anonymizer server.

4.4.3 Response Time Graph.



http request

Cloaking time of 100 mobile requests was computed as:
Cloaking time = endTime (CloakM (ms)) – startTime (CloakM (ms) as shown by (2).
endTime (CloakM(ms))=6500ms
startTime(CloakM(ms)=3950ms
Cloakingtime=6100ms-3900ms Cloaking time=2200ms
1000 milliseconds make up one second.
2200ms/1000ms=2.2s.
It took 2 seconds to cloak 100 mobile request.

4.4.4 Conclusion and future work.

The proposed framework is used by both snapshot and continuous queries thus it is able to protect the user from correlation attack. The anonymizer is a performance bottleneck and therefore it should be replicated and deployed behind a reverse proxy server to avoid a single point of failure. In the future, we plan to address more attack scenarios, such as attacks based on user preferences. Assume that each user is interested in certain types of queries, e.g., hotels, restaurants, bar etc. An attacker may use the additional knowledge to infer the query source.

5.0 References

[1] Beresford A. R.& Frank .S(2004). Mix Zones: User Privacy in Location-Aware Services. Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on. IEEE, 2.

[2] S. Gambs, O.Heen&C. Potin (2011). A comparative privacy analysis of geosocial networks. In Proceedings of the 4th ACMSIGSPATIAL International Workshop on Security and Privacy in GIS and LBS, SPRINGL ’11, pages 33–40, New York, NY, USA, ACM.

[3] .Mokbel, C. Chow,&W. G. Aref(2006.). The new Casper: Query Processing for location services without compromising privacy. In VLDB, pages 763–774.

[4] Mouratidis, k. & Yiu, M. L. (2010). Anonymous Query Processing in Road Networks. IEEE TKDE 22, 1,2–15.

[5] Kalnis, P., ghinita, G., mouratidis, k& papadias, (2007). Preventing Location-Based Identity Inference in Anonymous Spatial Queries. IEEE TKDE 19, 12, 1719–1733.

[6] Ghinita, P. Kalnis, and S. Skiadopoulos(, 2007). “PRIV´ E: Anonymous Location-Based Queries in Distributed Mobile Systems,” in WWW, pp. 371–380. [18])A Mobile Peer-to PeerSystem for Anonymous Location-Based Queries,” in SSTD, pp. 758–769

[7] Hashem, T., Kulik, L., Zhang, R(March 2010). Privacy preserving group nearest neighbor queries. In: Proceedings of the 13th International Conference on Extending Database Technology -EDBT ’10, New York, New York, USA, and ACM Press 489–500.

[8] Gkoulalas.V& mokbel,.F. (2009). Identifying unsafe routes for network-based trajectory privacy. In SDM pp. 942–953.

[9] Davison, M. R., Clarke, R., Smith, H. J., Langford, D., and Kuo, F Y(2003).“Information Privacy in a Globally Networked Society: Implications for IS Research,” Communications of the Association for Information Systems (12), pp. 341-365.

[10] Duckham, M. and Kulik, L. (2005). A formal model of obfuscation and negotiation for location privacy. In Proc. of the International Conference on Pervasive Computing.

[11] Olumofin Femi, K. Tysowski Piotr, Goldberg Ian, and Hengartner Urs(2010). Achieving efficient query privacy for location based services. In Privacy Enhancement Technologies (PETS),.

[12] Wei-Shinn Ku, Roger Zimmermann, Wen-Chih Peng, and Sushama Shroff(2007.)Privacy protected query processing on spatial networks. In ICDE Workshops, pages 215–220.

[13] Hu.H J. Xu, Q. Chen, and Z. Yang (2012). Authenticating location-based services without compromising location privacy. In Proc. SIGMOD, pages 301.

[14] Gkoulalas-divanis, a., verykios, v. S., and mokbel, M. F (2009) Identifying unsafe Routes for network-based trajectory privacy.

[15] Claudio Bettini, X. Sean Wang,&Sushil Jajodia(2005.). Protecting privacy against Location-based personal identification. In In 2nd VLDB Workshop SDM, pages185–199,

[16] Brinkhoff. K (2002,) “A Framework for Generating Network-Based Moving Objects,” (GeoInformatica, Vol.6 No.2, pp.153.

[17] Pingley.A W. Yu, N. Zhang, X. Fu, W. Zhao(2012). A context-aware scheme for privacy-preserving location-based services, Computer Networks, (56), 2551-2568.

[18] Pierangela.S(, 2001.) Pirotecting respondents’ identities in microdata release. IEEE Transactions on Knowledge and Data Engineering, 13:1010–1027.

[19] Marco Gruteser and Dirk Grunwal(2003). Anonymous usage of location-based services through spatial and temporal cloaking. In MobiSys ’03: Proceedings of the 1st international conference on Mobile systems, applications and services, pages.

[20] Man Yiu, Christian Jensen, Xuegang Huang, Hua Lu(2008). SpaceTwist: Managing the TradeOffs Among Location Privacy, Query Performance, and Query Accuracy in Mobile Services. 24th International Conference on Data Engineering,

[21] Hao.T Wen Peng &Wang Lee (2007) .Protecting Moving Trajectories Using Dummies. International Workshop on Privacy Aware Location BasedMobile Services.

[22] Ghinita. G, Panos Kalnis, Ali Khoshgozaran, Cyrus Shahabi & Kian Tan (2008). Private queries in Location Based Services: Anonymizers are not necessary. SIGMOD, 200831–42, New York, NY, USA,. ACM.

[23] Stenneth.L& Philip S. (2010.) Global Privacy and Transportation Mode Homogeneity Anonymization in Location Based Mobile Systems with Continuous Queries. 6th International Conference on Collaborative Computing: Networking, Applications and Work Sharing.

[24] Fuyu Liu, Kien Hua, Ying Cai(2009).Query l- Diveristy in Location- Based Services. International Conference On Mobile Data Management.

[25] Kido.H Y. Yanagisawa and T. Satoh (2005). Protection of Location Privacy using Dummies for Location-based Services. In ICPS

[26] Gisli R. Hjaltason , Hanan Samet, (1999)Ranking in Spatial Databases, ... Visual Information and Information Systems, p.317-324, June 02-04.

[27] Ashwin machanavajjhala, johannes gehrke, daniel kifer,muthuramakrishnanvenkitasubramaniam?(, 2006.)“l- diversity: privacy beyond kanonymity.icde