K-Anonymity: A Big Player in Location Privacy for Mobile Query Processing

Kshitija R. Patil

Abstract — With the rapid development in technologies, the use of mobile devices is growing continuously. In this modern era, the number of small devices like PDAs, smart phone and sensors with different kind of data are being used for various purposes. It is very essential to access the computing and data from static as well as the moving devices. Mobile computing environment encompasses the mobility of hardware, communication among the moving devices. Mobile computing also deals with access to data as well as software residing on the mobile devices. We can access the data from mobile hosts using various kinds of queries like location queries, existence queries and data queries [1]. Data can be accessed by processing these different kinds of queries using a variety of methods, which have been proposed by the researchers. While processing the queries and accessing data from the mobile devices or host, there is a danger of violating the location privacy of the devices, hosts. It is very necessary to hide their identity by applying some methods so that the privacy of the mobile nodes could be preserved. In this paper, we are discussing the concept of k-anonymity which is used by the researchers for proposing the privacy preserving methods for query processing in mobile computing environment. This paper takes a review of different strategies of query processing using the k-anonymity in mobile computing environment.

Key Words — LBS, anonymity, cloaking, location based query processing.

I. INTRODUCTION

Now a day the society is becoming more and more technology oriented as there is a vast variety of mobile devices available like PDAs, smart phones and sensors. Even in vehicles like modern cars, a number of small sized wireless devices are embedded in the form of computers, GPS navigators, or even multimedia centers. Therefore the moving vehicles can carry useful information which acts as a data sources for other vehicles. Mobile computing is known as a specialized class of distributed systems, where some terminals can move apart from joint distributed operations, move freely in the physical space and rejoins to a possibly different segment of a computer network in order to resume delayed activities [2]. The area of mobile computing is very broad, since there are multiple network connectivity options (3G, Wi-Fi, Bluetooth, UWB, etc.), with different features. While accessing data from the devices which may be moving or static, there is a need for preserving privacy of the mobile hosts, nodes. For example, the information about locality of places can be static or it also may be available on mobile devices. A mobile user looking for a restaurant can fire a query to find location. While accessing data regarding the location, there is a danger of malicious attacker that may collude with LBS provider to steal user’s location information and query logs causing harm to privacy of the node data.

The research is now targeted on privacy protection in location based query processing. After studying the different strategies of secure query processing, it is seen that the concept of k-anonymity model is used in most of the strategies.

II. LOCATION BASED QUERY PROCESSING IN MOBILE ENVIRONMENT

Location is a vital piece of information that relates to mobility. Mobility is the main feature of mobile environment. For accessing the Location based services, processing of spatial queries is getting more significance in mobile computing environment. GPS enabled mobile devices supports spatial query processing. Spatial queries are mobile queries that operate on the location information of mobile devices. Spatial queries are classified as: Location dependent query and Location aware query. In spatial queries, the location information is specified in the query or query is used to retrieve location information of mobile device. For accessing data location based services are used.

LBS can be defined as the specialized, multi-tiered, component web GIS (Geographical Information System) applications which can be invoked, published and located across the wired/wireless Web [3]. The location of the mobile devices can be determined by using GPS or mobile network triangulation. Such devices can report their locations to the LBS server through a wireless interface, or their locations can be obtained through ground-based radars or satellites. The important feature of LBS is to access location dependent data.

Figure 1 shows a general example of LBS. The working of LBS is as follows
1. An LBS user obtains the true position data of a user using a positioning device such as GPS.
2. The user sends the position data to a service provider.

Fig. 1: Working of LBS.
3. The service provider creates a reply message that responds to the received position data and sends it to the user.
4. The user receives a reply message [4].

III. PRIVACY IN LOCATION-BASED QUERY PROCESSING

There are two privacy issues in location-dependent queries:
i) The user must hide his location.
ii) The user must hide his identity (e.g., username, IP address, etc).
LBSs exploit knowledge about where users are located. The wide deployment of LBS systems faces the big challenge, the privacy-preserving management of location-based data. Without safeguards, extensive use of location-based services endangers location privacy of mobile users and exhibits significant vulnerabilities for abuse. Beresford et. al. defined location privacy as “the ability to prevent other parties from learning one’s current or past location” [5]. To hide the identity of a mobile user/node, it is necessary to make the node anonymous.

IV. WHAT IS K-ANONYMITY

Anonymity can be defined as “a state of being not identifiable within a set of subjects, the anonymity set” [6]. The concept of k-anonymity was originally introduced in the context of relational data privacy research. It pointed out the question of “How can a data holder release a version of its private data with scientific guarantees that the individuals who are the subjects of the data cannot be re-identified while the data remain practically useful” [7].

V. K-ANONYMITY FOR LOCATION PRIVACY

There are a number of approaches in the literature to solve the problem of privacy protection with location-based services, including:
- Cloaking;
- Generation of dummies;
- Private information retrieval (PIR).

Most of the authors have based their work on K-anonymity. The concept of k-anonymity depends on hiding the user’s location among K-1 neighbors. It requires a permanent communication and remote monitoring of the users. The robustness of these approaches depends totally on having number of neighbors at the time of receiving the requests. Some of the methods presented by different researchers for privacy preservation are:

A. Anonymous usage of location based services through spatial and temporal cloaking

Gruteser and Grunwald had used the concept of k-anonymity to measure microscopic location privacy for the first time [8]. In this, each query sent to the LBS is equivalent to one entry in a database, and the location-time information in the query serves as the quasi-identifier. It includes the user’s pseudonym, his position and the query time. In order to protect a user’s location privacy using k-anonymity, each of his queries must be indistinguishable from that of at least k-1 other user. The pseudonyms of these k users are removed from their queries, and the location-time pair in their queries is obfuscated to the same location-area and time-window, large enough to contain the users’ actual locations. It allows a wide range of users to benefit from the location privacy protection with personalized privacy requirements. A novel spatio-temporal cloaking algorithm, called CliqueCloak provides location k-anonymity for mobile users of a LBS provider. The cloaking algorithm is run by the location protection broker on a trusted server. It anonymizes messages from the mobile nodes by cloaking the location information contained in the messages to reduce or avoid privacy threats before forwarding them to the LBS providers. The anonymization algorithms on Java server platform evaluates them using automotive traffic simulations based on US geological survey (USGS) cartographic material.

Fig.2 Basic k-anonymity model

B. A Customizable k-Anonymity Model for Protecting Location Privacy

A customizable k-anonymity model for protecting privacy of location data has two unique features. It provides a customizable framework to support k-anonymity with variable k. It allows a wide range of users to benefit from the location privacy protection with personalized privacy requirements. Second, a novel spatio-temporal cloaking algorithm, called CliqueCloak which provides location k-anonymity for mobile users of a LBS provider. The cloaking algorithm is run by the location protection broker on a trusted server. It anonymizes messages from the mobile nodes by cloaking the location information contained in the messages to reduce or avoid privacy threats before forwarding them to the LBS providers. The model enables each message sent from a mobile node to specify the desired level of anonymity as well as the maximum temporal and spatial tolerances for maintaining the required anonymity. [9].

C. The New Casper

Casper[10] presented by M.F. Mokbel et.al., is a new framework in which mobile and stationary users can entertain location-based services without revealing their location information. Casper framework consists of two main components, the location anonymizer to hide the identity of mobile node and the privacy-aware query processor. The location anonymizer blurs the users exact location information into cloaked spatial regions based on user specified privacy requirements. To deal with the cloaked spatial areas, the privacy-aware query processor is embedded inside the location-based database server which serves the data.
D. Anonymous Communication Technique using Dummies for Location-based Services

A new anonymous communication technique protects location privacy using anonymity Set. The technique makes the use of dummies to protect the location privacy of people using Location Based Services [4]. In this technique, a user sends true position data with several false position data (‘dummies’) to a service provider. Then, it creates a reply message for each received position data. The user simply extracts the necessary information from the reply message. In this way, even if the set of position data is stored by the service provider, it cannot distinguish the true position data from the set of location data. The technique uses two dummy generation algorithms to prevent service providers from finding the true position data. They are:

1) Moving in a Neighborhood (MN)
2) Moving in a Limited Neighborhood (MLN)

E. Privacy Protected Query Processing on Spatial Networks

Spatial query processing deals with processing of location information of mobile devices. A K-anonymity mechanism is proposed to preserve user privacy on Spatial Networks [11]. Two novel query algorithms, PSNN and PSRQ have been presented for answering nearest neighbor queries and range queries on spatial networks without revealing private information of the query initiator. The main theme of the algorithms is to hide the exact mobile user location with a cloaked region. The cloaked region covers the query requester and at least K – 1 other users based on the K-anonymity concept. The spatial queries are executed based on both the cloaked region and the underlying networks. A candidate result set will be returned to the requesting user who filters out the exact answer.

F. Protecting Location Privacy with Personalized k-Anonymity

Personalized k-Anonymity is a scalable architecture for protecting the location privacy from various privacy threats resulting from uncontrolled usage of LBSs. This architecture includes a personalized location anonymization model and a suite of location perturbation algorithms. A flexible privacy personalization framework is used to support Location k-anonymity for a wide range of mobile clients with context-sensitive privacy requirements. This framework enables each mobile client to specify the minimum level of anonymity that it desires and the maximum temporal and spatial tolerances that it is willing to accept when requesting k-anonymity-preserving LBSs. An efficient message perturbation engine implements the location privacy framework. The anonymity server performs location anonymization on LBS request messages of mobile clients such as identity removal and spatio-temporal cloaking of the location information [12].

G. Protecting Privacy in Location-based Services Using K-anonymity without Cloaked Region

Traditional K-anonymity method needs complex query processing algorithms at the server side. Another approach SpaceTwist [14] rectifies the shortcoming of traditional K-anonymity system by using the incremental nearest neighbor (INN) queries processing techniques at the server side. However, SpaceTwist does not guarantee K-anonymity so it may fail at some critical instance. A new framework called KAWCR is a K-anonymity Without Cloaked Region model which rectifies the shortcomings of simple k-anonymity and spacetime model and retains the advantages of k-anonymity as well as spacetime. KAWCR only needs the server to process INN queries and can guarantee that the users issuing the query is indistinguishable from at least K-1 other users. The communication cost of KAWCR for KNN queries is lower than that of both traditional K-anonymity and SpaceTwist on some datasets.

VI. ANALYSIS

After studying a number of methods for secure query processing it is seen that the model based on spatio-temporal cloaking algorithm[8] enables each message sent from a mobile node to specify the desired level of anonymity as well as the maximum temporal and spatial tolerances for maintaining the required anonymity. The location k-anonymity model with multi-dimensional cloaking and tunable k parameter can achieve high guarantee of k anonymity. It provides high flexibility to location privacy threats without significant performance penalty [9]. The
Casper achieves high quality location-based services while providing anonymity for both data and queries [10]. The technique of sending true position data with several false position data (‘dummies’) [4] to a service provider protects location privacy using Anonymity Set and can be applied in practical LBSs. The query processing solutions discussed by most of the researchers are based on Euclidean metrics. But in real life, mobile users cannot move freely in space because they are usually constrained by underlying networks (e.g., cars on roads, trains on tracks, etc.). Therefore, the K-anonymity mechanism with PSNN and PSRQ algorithms [11] proves better for spatial networks. By using realistic location data that is real road maps and traffic volume data the effectiveness of location cloaking algorithms is tested. It shows that the personalized location k-anonymity model [12], together with location perturbation engine, can achieve high resilience to location privacy threats without introducing any significant performance penalty. As Compared with traditional k-anonymity models, KAWCR[13] can guarantee that the user issuing the query is indistinguishable from at least K-1 other users. KAWCR only needs INN query processing algorithm while traditional K-anonymity needs complex processing algorithm at the server side.

**CONCLUSION**

It is seen that strong protection for user information can be attained, if the server is made capable of retrieving location related information without being aware of the user’s position or the point of interests he/she is requesting. The anonymity and cloaking-based approaches used to address the problem of location privacy cannot provide stringent privacy guarantees without incurring costly computation and communication overhead. These methods require a trusted intermediate anonymizer to protect a user's location information during query processing. All the privacy preserving schemes for location data or the nodal identity require to hide their identity by anonymising them with k-1 users/nodes. Hence k-anonymity plays a very important role in location privacy. The study of the various methods and algorithms shows that the location k-anonymity model with multi-dimensional cloaking and tunable k parameter can achieve high guarantee of k anonymity and high flexibility to location privacy threats without significant performance penalty. Due to its simplicity, the k-anonymity scheme for location privacy has become very popular.

**REFERENCES**


**AUTHOR’S PROFILE**

Kshitija R. Patil
M. Sc.(Comp. Sc.), B.Ed., M.Phil(Comp Sc.),
Asst. Prof., Degree College of Physical Education, H.V.P.M., Amravati (Mah.)

Dr.Vinod Patil
M.Sc.,Ph.D.,Member, Board of Studies in Computer Science, S.G.B.Amravati University,Amravati.

National Conference on “Advanced Technologies in Computing and Networking” ATCOn-2015
Special Issue of International Journal of Electronics, Communication & Soft Computing Science and Engineering, ISSN: 2277-9477

307