Computer Networks 140 (2018) 1–14

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Modeling privacy approaches for location based services

Pratima Biswas[a], Ashok Singh Sairam[b,*]

[a] Indian Institute of Technology Patna, Patna 801106, India
[b] Indian Institute of Technology Guwahati, Guwahati 781039, India

ARTICLE INFO

Article history:
Received 22 September 2017
Revised 13 February 2018
Accepted 28 April 2018
Available online 1 May 2018

Keywords:
Location based services
Privacy
k-anonymity
Queuing theory

ABSTRACT

Locationbased services (LBS) use geospatial data of mobile device to provide information in real time. A key concern in using these services is the need to reveal the user's exact location, which may allow an adversary to infer private information about the user. To address the privacy concerns of LBS users, a number of security approaches have been proposed based on the concept of *k*-anonymity. The central idea in location *k*-anonymity is to find a set of *k*-1 users close to the actual user, such that the locations of these *k* users are indistinguishable from one another, thus protecting the identity of the user. A number of performance parameters like success rate, amount of privacy achieved are used to measure the performance of the *k*-anonymity approaches. However, there is no formal model to compare the different *k*-anonymity approaches. Moreover, these proposals also make the implicit, unrealistic assumption that the $k-1$ users are readily available. Thus they ignore the *turnaround* time to process a user request, which is crucial for a real-time application like LBS. In this work, we model the *k*-anonymity approaches using queuing theory to compute the average sojourn time of users and queue length of the system. To demonstrate that queuing theory can be used to model all *k*-anonymity approaches, we quantitatively compare three different *k*-anonymity approaches with varying degree of complexity - top-up, bottom-down and bulk processing. The proposed analytical model is further validated with experimental results.

1. Introduction

With the wide availability of location-aware devices and advancement of positioning technologies like Global Positioning Systems (GPS) to determine exact locations of users and objects of interest, a new class of applications called locationbased services (LBS) have become highly popular. These applications can vary from utility applications like finding points of interest, friends currently present in ones vicinity to serious applications like sending alarm messages during emergency etc. One of the main concern in using these services, is that they require revealing ones location which may allow an adversary to infer sensitive information about the user. To address the privacy concerns of LBS, a number of approaches have been proposed, popular among them are those approaches that implement the concept of *k*-anonymity. In this approach, the key idea is to find a set of *k* users confined in a given geographical area such that they are indistinguishable from one another, thus protecting the identity of the user.

Central to the idea of *k*-anonymity in LBS is a trusted third party (TTP) which is delegated with the task of anonymization. When a LBS query arrives at the TTP, it finds $k-1$ other users in the vicinity of the user and sends the obfuscation area to the LBS server. This is known as *cloaking* or position *obfuscation*. The different LBS privacy approaches using *k*-anonymity basically differ in the way how the $k-1$ other users are selected. These approaches make an unrealistic, implicit assumption that the $k-1$ other users are readily available. However, in practice queries for anonymization will arrive at unpredictable times and when they arrive other users may not be available. For example in Clique–Cloak [1] approach, whenever a query arrives it is checked if the location point of the user forms a clique with $k-1$ other users. In such a case, the first $k-1$ users will always have to wait. Thus the natural question that arise is how long a LBS query may have to wait before it can be served, for how long will the TTP be busy in computation and so on. The existing *k*-anonymity approaches consider different performance parameters like success rate, amount of privacy level achieved, etc. but do not consider parameters like average response time of a query which is very important as the queries are fired in real time and users want fast response.

Although the idea of *k*-anonymity has been well established [1,2] and effectively deployed to solve the upcoming challenges of location privacy [3,4], no mathematical model are available to evaluate them. The contribution of our work can be summarized as follows:

* Corresponding author.
E-mail addresses: pratima.pcs13@iitp.ac.in (P. Biswas), ashok@iitg.ernet.in (A.S. Sairam).

https://doi.org/10.1016/j.comnet.2018.04.016
1389-1286/

(i) Modeling of k-anonymity privacy approaches: Our first aim is to develop a mathematical framework that can be used to evaluate k-anonymity privacy approaches in terms of request-to-response time of a query, the number of queries present in the TTP, length of a busy period and length of an idle period. Next we adapt this framework to model specific privacy approaches available in literature.
(ii) Experimental validation of the model: The second goal of our work is to experimentally compute the performance parameters estimated using our mathematical model and compare the results.
(iii) Comparative analysis: The last objective is to compare the various k-anonymity approaches available in literature based on our proposed model. The result of our analysis show that the bottom-up privacy approach clearly outperforms the top-down approach. Bulk processing privacy approaches have the nice proper of generating cloaking area with high spatial resolutions, however, they have higher turnaround time and low anonymization probability rate.

In this work, we use the concept of single service systems and bulk service systems of *Queueing theory* to model the k-anonymity LBS privacy approaches that uses a TTP. In order to demonstrate our approach, we apply the concept to some well-known k-anonymity LBS privacy approaches [1,5–7]. Results show that our mathematical model as compared to experimental results has a high accuracy with an error percentage of about 2.5%.

The remainder of our work is organized as follows. Section 2 reviews state-of-the-art in location services, privacy threats and defense mechanisms. The background of our work is presented in Section 3. In Section 4, we describe how queuing theory can be used to model the k-anonymity privacy approaches. The proposed queuing model is given in Section 5. Experimental results are given in Section 6 and finally the concluding remarks are given in Section 7.

2. Prior art

Location based services make use of a user's location information to provide personalized information service to the user. There are many application scenarios of LBS like navigation services, emergency services etc. that have enjoyed great commercial success. However, many users fear that the location information may be misused to intrude into their privacy, like determine their political inclination, health problems, friends and acquaintances, etc. The techniques proposed to preserve the privacy of a user primarily concentrate on protecting three user attributes - user identity, spatial information and temporal information. The user identity can be protected using pseudonyms but an attacker can infer the user's identity by evaluating the location information known as *indirect location privacy problem* [8]. A user may protect his position information by providing coarse position information, a technique known as *obfuscation* or *cloaking* [9]. The main issue with this approach is to minimize the cloaking area for a given privacy requirement without deteriorating the quality of service. Temporal information is related to the real-time updates of a user's position. A possible protection mechanism for temporal information is delayed updates but there are certain services (eg. navigation system) which require real time update. Moreover, if the temporal information is omitted it can be inferred by understanding when the location update was triggered. Thus among the user attributes, protecting spatial information is the most crucial for location privacy.

Various techniques have been proposed in literature to protect the user privacy in LBS applications [10,11]. These methods mainly comprise using encryption techniques [12,13], k-anonymity [1,5,6], mix zones [14,15], machine learning [16,17] and many others. Among these approaches, k-anonymity has been extensively used and it has proved its effectiveness [1,2]. k-anonymity is a general privacy concept which ensures that in a set of k objects, the target object is indistinguishable from the other $k-1$ objects. A description of the k-anonymity approaches used in location privacy is presented in Section 4.1.

Selecting the right performance metrics is crucial to designing a privacy protection algorithm. Basically the hypothesis is how correctly an attacker can infer the user's location and robustness of the approach. Entropy [18] is used to measure the uncertainty in identifying a user's real location. A classical privacy protection metric is measuring the spatial and temporal cloaking based on k-anonymity metric. Privacy preserving methods cannot assume a binary goal; the adversary either has or has not learned something sensitive. A number of attacks are possible to infer the sensitive with a certain degree of confidence. The l-diversity measure [19] ensures that there are l distinct values for the sensitive attribute in each equivalence class. However, a l-diverse equivalence class may have a value appear more frequently than others, thus enabling the adversary to make an inference. This led to the introduction of entropy in l-diversity.

Li et al. [20] showed that l-diversity is difficult to achieve and it is not sufficient to prevent attribute disclosure. They introduced a performance metric t-closeness. The metric tries to minimize the inference that can be drawn from the distribution of a sensitive attribute in a class by ensuring that the distance of the distribution in a class is no more than the distribution of the attribute in the whole table by a threshold t. Shokri et al. [21] provide a formal framework to model the adversary. The idea behind modeling the adversary is to determine whether the adversary is able to correctly find the intended answer. The privacy is quantified by measuring the distance between the output of the attack and the actual answer, which they call as the *correctness* of the attack. However, this work do not provide a mechanism to quantify and compare the available k-anonymity techniques.

The most widely used privacy measure is k-anonymity with spatial cloaking. In an LBS using k-anonymity approach, spatial cloaking is inevitable. The k-anonymity requirement ensures that a user is indistinguishable from $k-1$ other users. In order to ensure k-anonymity, the user's exact location is not used in the query but a larger spatial region (A) covering the entire k users is employed. The spatial cloaking ensures that a user's exact location is obscured into a spatial region A (also called cloaking region), such that the probability of locating the user within the spatial region is $\frac{1}{A}$. A major issue with spatial cloaking is that the quality of service is degraded due to the larger spatial region used in processing the query. Thus the main objective in LBS using k-anonymity is to optimize the cloaking region. Spatial query processing techniques [22,23] for a cloaking region has been proposed but they are complex and their efficiency is not known.

Although there is a large body of work on k-anonymity for LBS, the lack of a mathematical model to evaluate existing as well as new proposals have resulted in difficulty in practical adoption of these proposals. To the best of our knowledge, a generic framework or model is not available for their comparative analysis. The only work that model location privacy is the work by Huang et al. [24]. However, their work is based on modeling eavesdropping in wireless networks with the objective to compromise their location. The aim of our work is to provide a model for analysis of k-anonymity techniques used in LBS for privacy preservation.

3. Location based services

LBSs [25] are information services that exploit a mobile user's current location to provide value added information. The basic