User Density and Spatial Cloaking Algorithm Selection:
Improving Privacy Protection of Mobile Users

Matthew Chan
Department of Computer Information Systems
Borough of Manhattan Community College, CUNY
New York, NY 10007, USA
machten@bmcc.cuny.edu

Hassan Elsherbini, Xiaowen Zhang
Department of Computer Science
College of Staten Island, CUNY
Staten Island, NY 10314, USA
hassan.elsherbini@cti.csi.cuny.edu
xiaowen.zhang@csi.cuny.edu

Abstract—Data sharing and privacy protection of mobile users have always been a challenge to research and development, as well as commercial and enterprise deployment of the Global Positioning System (GPS) and location-based mobile applications. The concepts of k-anonymity, two spatial cloaking algorithms—Nearest Neighbor Cloak (NNC) and Hilbert Cloak (HC)—that utilize k-anonymity and Hilbert cloak, which utilize the concept of k-anonymity to generate the k-ASR in an anonymizer.

Keywords—user density; privacy; spatial cloaking; location-based services; k-anonymity; mobile technology

I. INTRODUCTION

Nowadays smart phones and other mobile devices are becoming increasingly ubiquitous. These mobile devices with positioning capabilities utilize technologies such as the Global Positioning System (GPS) that enables and facilitates Location Based Services (LBS). In these services the need to protect user location privacy and user confidentiality is a necessity. One of the techniques used to satisfy such necessity is spatial cloaking which is the process of anonymizing a user’s location to a degree in which the likelihood of inferring it is very low. This leads to the topic of k-anonymity, a highly-adopted concept used in spatial cloaking. Location obfuscation [1] can be used to obscure and conceal the actual query location of a user by expanding his coordinates to a larger space. We plan to implement the k-anonymity mechanism in an R*-tree [2] indexed database and to conduct experiments with an attempt to confirm whether the proposed system can achieve improved performance. The proposed research seeks to adopt an adaptive scheme that utilizes cloaking algorithms such as Nearest Neighbor and Hilbert cloak [3] to improve the security, privacy protection and performance of a system. Other approaches [4-7] also use k-anonymity to provide location privacy to mobile device users seeking LBS services.

II. K-ANONYMITY

Suppression and generalization are some of the methods commonly used to hide attributes that may reveal privacy when archiving or sharing data sets. k-anonymity is one of the most adopted methods used to tackle the issues of protecting individual privacy while sharing the data, and at the same time, maintaining the usefulness and accuracy of the data. Introduced by L. Sweeney, k-anonymity attempts to protect individual privacy by adding k-1 records to the data set to reduce the risks of the attributes of original record being reidentified [8]. This method expands the original record to a set of records with similar attributes. In other words, k-anonymity means the original record is mixed with k-1 records, increasing the difficulty of the original record being identified. While k-anonymity does provide a solution to prevent sensitive data disclosure, it remains impractical when dealing with high-dimensional relations. Aggarwal [9] discusses the evident problem with dimensionality and how anonymizing a high-dimensional relation leads to unacceptable loss of information.

In LBS k-anonymity is widely adopted to prevent identity compromise via location querying. An example of a framework that utilizes k-anonymity for such task is as follows (see Fig. 1). First, a user sends his query and location information to an anonymizer, a secured server which serves as a middle man between the user and the LBS provider. In some cases (e.g. P2P network), an anonymizer could be part of user device. Next, the anonymizer removes the id (a unique identity number) of the user and constructs a k-ASR (anonymized spatial region), a minimum bounding rectangle or circle encompassing the location of the query issuer with k-1 other user locations, to cloak the user’s location. Finally, the generated k-ASR is sent to the LBS provider, which processes and returns a set of candidates. The anonymizer then removes the false hits from the set and forwards the results to the user.

III. SPATIAL CLOAKING ALGORITHMS

Two spatial cloaking algorithms are discussed: Nearest Neighbor cloak and Hilbert cloak, which utilize the concept of k-anonymity to generate k-ASR in an anonymizer.
A. Nearest Neighbor Cloak

When a user $U$ initiates a query, NNC computes the set $S_0$ which will include user $U$’s $k$-1 closest users (Fig. 2). Next, a random user $U_i$ is chosen from the set $S_0$ and then a new set $S_1$ is computed to contain $U_i$ and his closest $k$-1 users. The final set $S_k$ will be the union set of user $U$ and the set $S_1$. Hence, the $k$-ASR becomes the MBR (minimum bounding rectangle) that covers all the users of the anonymized set $S_k$. Due to the randomness of the algorithm, the probability that user $U$ is at or close to the center of the $k$-ASR is remote. Thus, NNC is not vulnerable to the center-of-ASR attack which other cloaking algorithms such as Center Cloak are exposed to.

![Fig. 2. NNC and sets $S_0$, $S_1$, and $S_k$.](image)

B. Hilbert Cloak

First, with a query from a user $u$ and also the anonymity value $k$, HC computes and sorts the Hilbert curve values of all the users. It then separates the users into buckets in which each bucket has exactly $k$ users (except the last bucket which may have up to 2$k$-1 users so that no bucket will have less than $k$ users). That is, HC partitions the user population into many $k$-ASRs by obtaining the Hilbert value $H(u)$ of every user. This is achieved by employing the Hilbert space-filling curve [10]. The curve converts the two dimensional coordinates of all users into single dimensional Hilbert values. The conversion guarantees with a high probability that any two locations will remain in nearby vicinity in the one dimensional space if they were in close vicinity in the two dimensional region. Finally, given $H(u)$ and rank$(u)$, HC identifies the $k$-bucket containing rank$(u)$ and then the MBR, which encompasses the $k$ users in the bucket, becomes the $k$-ASR. Using rank$(u)$, the initial position as well as the end position establishing the $k$-bucket can be computed as the following:

\[\text{initial} = \text{rank}(u) - (\text{rank}(u) \mod k), \quad \text{end} = \text{initial} + k - 1\]

Compared to NNC, HC always guarantees $k$-anonymity. This is due to the fact that HC satisfies reciprocity.

IV. EXPERIMENT RESULTS

Assuming NNC and HC behave significantly differently depending on the user density, it would then be advantageous to select a different cloaking method for users with different density. We conducted experiments to confirm the idea. The dataset used for the experiments was based on California in the U.S. The number of users used was one third of the number of points of interest (POI), randomly chosen to be within ten meters from the location of a POI. We first computed the user density for each user. Then we isolated the one thousand users with the highest density and another thousand users with the lowest density. The user density was based on the total number of other users that were within a three kilometer circle from a user. Using NNC and HC, we computed the size of the $k$-ASR and also the number of POI that were located inside the $k$-ASR for both the high and low density users. The $k$ used was eighty as it provided a sufficient amount of anonymity for most users. The results (see Fig. 3) show that the NNC has a smaller ASR area than that of HC for both high and low user density. The average ASR of NNC for high density users is 19.25 sq km and 1838.17 sq km for the low density users. On the other hand, the average ASR of HC for the high density users is 108.89 sq km and 3322.65 sq km for the low density users. Since NNC performs better than HC in both high and low user density, our current experimental outcomes illustrates that user density does not provide a beneficial indicator for selection of a cloaking method.

![Fig. 3. Preliminary experiment results.](image)

REFERENCES


