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

Matthew Chan, Computer Info. Sys. Dept, Boro of Manhattan Comm. College, CUNY
Hassan Elsherbini, Department of Computer Science, College of Staten Island, CUNY
Xiaowen Zhang, Department of Computer Science, College of Staten Island, CUNY

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, as well as user density's impacts on the performance are discussed in this research . The proposed research seeks to examine and adopt an adaptive scheme that utilizes these cloaking algorithms to improve security, privacy protection and performance of a system, using *k*-anonymity to generate the *k*-ASR in an anonymizer.

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 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 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 to improve the security, privacy protection and performance of a system.

K-ANONYMITY

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. In LBS *k*-anonymity is widely adopted to prevent identity compromise via location querying. 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 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.

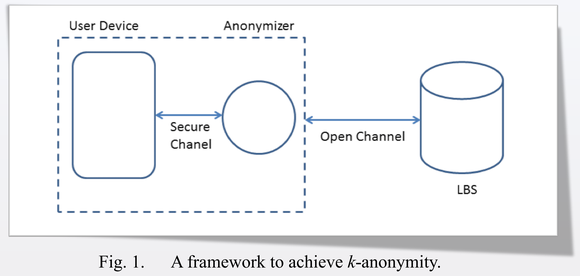

Fig. 1. A framework to achieve *k*-anonymity.

SPATIAL CLOAKING ALGORITHMS

Nearest Neighbor Cloak (NNC): Utilizing *k*-anonymity, NNC is one of the algorithms used by an anonymizer for spatial cloaking. 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_2 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_2. 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.

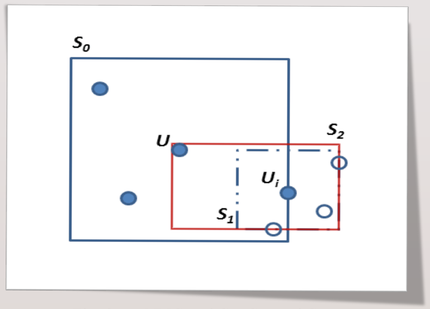

Fig. 2. NNC and S_0 , S_1 , and S_2.

Hilbert Cloak (HC): Hilbert cloak, another spatial cloaking algorithm, utilizes the concept of *k*-anonymity to generate *k*-ASR in an anonymizer. 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. 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:

$$initial = rank(u) - (rank(u) \bmod k), \quad end = initial + k - 1$$

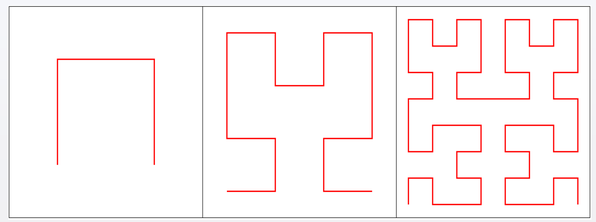

Fig. 3. Hilbert curves of first, second and third order.

EXPERIMENTAL 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. 4) 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.

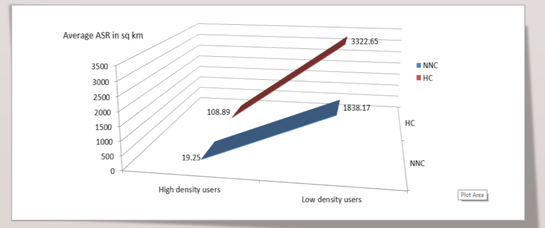

Fig. 4. Preliminary experiment results.

REFERENCES

S. Papadopoulos, S. Bakiras, and D. Papadias, "Nearest neighbor search with strong location privacy," in Proceedings of the VLDB (Very Large Data Bases) Endowment, 3(1-2), pp. 619-629, 2010.

A. Pingley, W. Yu, N. Zhang, X. Fu, and W. Zhao, "Cap: A context-aware privacy protection system for location-based services," in Distributed Computing Systems, ICDCS'09, 29th IEEE International Conference, pp. 49-57, 2009.

B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, "Achieving k-anonymity in privacy-aware location-based services," in IEEE INFOCOM 2014-IEEE Conference on Computer Communications, pp. 754-762, 2014.

L. Sweeney, "k-anonymity: a model for protecting privacy," in International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, vol. 10, no. 5, pp. 557–570, 2002.

C. C. Aggarwal, "On k-Anonymity and the curse of dimensionality," in Proceeding of Very Large Data Bases (VLDB), pp. 901–909, 2005.

M. F. Mokbel, W. G. Aref, and I. Kamel, "Analysis of multi-dimensional space-filling curves," in GeoInformatica, vol. 7, no. 3, pp. 179–209, 2003.