**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?**

A release provides k-anonymity protection if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release.

**Attack : Re-identification by linking**

For convenience, I focus on person-specific data, so the entities are people, and the property to be protected is the **identity of the subjects** whose information is contained in the data. However, **other properties could also be protected.** The formal methods provided in this paper include the *k*-anonymity protection model.

so attributes which appear in private data and also appear in public data are candidates for linking; therefore, these attributes constitute the quasi-identifier and the disclosure of these attributes must be controlled. It is believed that these attributes can be easily identified by the data holder.

Sensitive information for ST data: frequent track, sensitive location. …

How to compare …..

Encryption

---speed

不同类型的数据集，主要是不同大小的，测试加密速度。

---visualization

视觉上直观的比较两种加密的不同。

---security

我现在做的是 对于location-k anonymity, 把每一个时间段的**任意位置**都进行加密，试图满足要求。MIX-ZONE 要求有一个区域，通过计算entropy来确定privacy的情况。K-area 把一个区域分成几个部分，每一个部分都已K个sensitive area。Uncertainty, 通过计算sensitive area 的面积与模糊后的面积的比值来确定privacy的情况。

--- Accuracy/usefulness: distance between real and fake locations. (time ?)

--- Security: the difficulty of being recognized.

如果以家的位置为sensitive information, k-anonymity取决于人的分布范围，如果分布紧密，也会暴露位置。

到底是保护人的identity还是sensitive location的位置？？？？

应该还是注重identity。

How to be recognized? Or the possibility of being recognized???

Track can be uniquely identified by 4 points. 我们可以观察经过这两种方式加密后，能被独特识别出来的路径有多少。

For k: 对于目标用户，加密其常走路径，得到结果。

For uncertainty:　对于相同的目标用户，加密其常走路径，得到结果。

Details:

The area of Anonymizer & uncertainty region is the same. Q

usersRoute = Preprocessed(userdata) % 每个user多条路径

% encryption

For each route of each user:

getEachPoint = usersRoute{I,1}{j,1}(k);

kanonymityUserRoute = adaptiveinterval(getEachPoint);

uncertaintyUserRoute = uncertainty(getEachpoint);

update;

% comparing

[a,t]= Fix-spatial-and-temporal resolution

percentOfBeingrecognizedTok=findSameRoute(kanonymityUserRoute,a,t);

percentOfBeingrecognizedToUncertanity = findSameRoute(uncertaintyUserRoute,a,t);

accuracy =

k-anonymity ------------------ ST data

2015.4.4

Clustering on horizontally partitioned data

UTM projection

1. **Hilbert Cloaking Algorithm**
2. **Dummy Generation Algorithm //**
3. **Location K-anonymity**
4. **K-anonymity customized**
5. **Mix-zone**

E MIX-ZONES

In this context,

we describe two metrics that we have developed for measuring location privacy, one based on anonymity sets and the other based on entropy. Finally, we move from theory to practice by applying our methods to a corpus of more than three million location sample points obtained from the Active Bat installation at AT&T Labs

The more challenging problem we explore in this articleis to develop techniques that let users

benefit from location-based applications while at the same time retaining their location privacy.

At an earlier stage in this research, we tried this kind of attack on real location data from the Active

Bat and found we could correctly deanonymize all users by correlating two simple checks: First, where does any given pseudonym spend most of its time? Second, who spends more time than anyoneelse at any given desk? Therefore, longterm pseudonyms cannot offer sufficient protection of location privacy.

UTILIZING SPATIO-TEMPORAL DATA INDEX

FOR LOCATION PRIVACY PROTECTION

The main idea of these

solutions is to obfuscate users’ location [9, 10], or to anonymize

location information [1, 8, 11, 12]. Among those solutions, there

are two limitations. Firstly, most of them deal with only spatial

obfuscation, not temporal one. Secondly, these algorithms are

separated from the database layer. This makes the algorithms go

through two-phase: retrieving the exact location of user from the

database (phase 1), and then obfuscating this information in the

algorithm (phase 2).

Attack: Query Sampling Attack

Existed algorithms:

1. Location Obfuscation

**D. M. Gruteser and D. Grunwald. Anonymous usage of locationbased**

**services through spatial and temporal cloaking. In**

**ACM/USENIX MobiSys, 2003.**

In the context of LBSs and mobile users, location *k*-anonymity

demands that location information contained in a message sent

from a mobile user to a LBS should be indistinguishable from

at least *k−*1 other messages from different mobile nodes

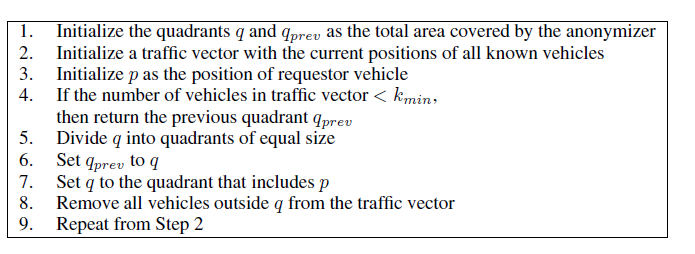
**Adaptive-Interval Cloaking Algorithms**

The key idea underlying this algorithm is that a given degree

of anonymity can be maintained in any location—

regardless of population density—by decreasing the accuracy

of the revealed spatial data.



**Definition:**

**Total covered area, A. k-anonymity, K=3. q, q\_prev.**

**Cell structure, each cell for each user.**

**A: 16**

**Stay zone 1km 10min**

**Two people are in the same zone if distance meet standard**

**And time is in the same**

**A:[t1,t2] B:[t3,t4] t2>t3 & t1<t3 &//// (t1-t4)<4(t2-t3)**

**Or t4>t1 & t3<t1 &//////(t2-t3)<4(t4-t1)**

**E: Our model has two unique features. First, we provide a customizable framework to support kanonymity with variable k, allowing a wide range of users to benefit from the location privacy protection with personalized privacy requirements. Second, we design and develop a novel spatio-temporal cloaking algorithm, called CliqueCloak, which provides location k-anonymity for mobile users of a LBS provider.**

Each message can specify a

different *k* anonymity value based on its specific privacy requirement.

Furthermore, each message can specify its preferred

*spatial and temporal tolerance* level in order to maintain

the desired *k* anonymity property.

*CliqueCloak*