(N,k)-Diverse anonymity：a strong privacy protection scheme for Location Based Services

Juncheng Pan, Hongli Zhang, Huimin Deng, Dongsheng Ding, Hongwei Li
School of Computer Science and Technology, Harbin Institute of Technology
Harbin, China
steven.pan.v@gmail.com, zhanghongli@hit.edu.cn, donsding@126.com

Abstract*—Today's location based service (LBS) places great demand on protecting the users' location related privacy. Most existing work adopts spatial k-Anonymity to ensure a specific query not be related to its user with a probability exceeding 1/k. However, it does not guarantee the privacy of user's *location attribute(LA)*, user's relationship with the location he located in. For example, if a spatial cloaked region is focused on the vicinity of a hospital, then location attribute of its user can be assure as patient or doctor. This disadvantage is result from traditional anonymity lack of diversity, which means the anonymity set(k users) is monotonous in terms of the total number of the type of its users' LA. To conquer this disadvantage, an enhanced privacy scheme, *(N, k)-Diverse anonymity* is proposed to prevent the user's location attribute from being leakage, through diversifying traditional spatial k-Anonymity. In order to achieve a diverse anonymity region, there are two steps: (1). Generate a diverse spatial cloaking region; (2) Render the region to be a k-anonymity region. Correspondingly, we design an N-Location diversity Cloaking Algorithm (N-LDC) to realize step1, and then propose Node Increment Cloaking (NIC) and Region Increment Cloaking (RIC), which are derived from traditional k-anonymity cloaking, to realize step 2. Further, we improve our N-LDC to secure-NLDC, to defend Central PS attack, which strengthen the robustness of (N,k)-Diverse anonymity. At last, our theoretical and experimental analysis suggest that our techniques are highly effective in guaranteeing the user's privacy in terms of time, space and robust to attacks.

***Keywords*—location base service, location attribute; location diversity; k-Anonymity; (N,k)-Diverse anonymity, privacy**

I. INTRODUCTION

In recent years, Location Based Service(LBS) have gained tremendous popularity in today's networked society. Typical LBSs are centralized location broker service[27], road hazard detection and prediction[12,13,27,28], navigation[12,13], resource scheduling and allocation[28], etc. Nevertheless, users' location privacy may be infringed from LBS. So it is very important to protect its users' location related privacy. For example, the popular spatial k-Anonymity[3,4,7,17,27] can ensure each reported location being an enlarged spatial region which contains at least k undistinguished users so that attackers can't map a specific query to its object with probability higher than 1/k. Although k-Anonymity model can guarantee the query privacy [26] of LBS's users, it fails to guarantee the privacy of user's *location attribute*, user's relationship with the position he located in. For example, if a spatial cloaked region which focus on the vicinity of a certain hospital, location related identity of its user would be exposed as a patient or a doctor, even though the query's source (who issues the query) can't be certain with probability higher than 1/k in one snapshot. We term this situation as user's location attribute leakage. User's location attribute can be served as quasi attribute [21, 29] to re-identify the user's identity by record linkage technologies [4]. User's location attribute leakage is due to the spatial cloaked region fails to protect his or her location attribute because the region may lack a key property, termed *location diversity*. A region is said to be *location diversity* if it covers parts of two or more PSs(public sites)' vicinity. The opposite is location homogeneity that is the cloaked region is the vicinity of one PS only.

For a better understanding the location diversity and homogeneity, an example of spatial cloaking is given by Tables 1, and 2 and Figue 1. Table 1 records several user's location information, their ID, public sites they are nearest to and user's *location attributes*. Table2 records information of two cloaked regions. The first cloaked region R1 in table2 covers A,B,C, and it focus on shoe factory only, attacker can deduce that these three users may be factory workers. The region R2 intersects with vicinities of three public sites, namely shoe factory, Middle school and restaurant. So there would be three location attribute candidates for the potential message issuer in the cloaking region. Due to two different number of location attributes, the R1 is location homogeneity while the R2 is location diversity.

To relieve user's location attribute leakage, a direct method is to increase the number of location attribute candidates for potential user. Based on this intuitive methodology, an enhanced privacy property ***(N,k)-Diverse anonymity is proposed***. The property means that if a spatial cloaked region satisfies the property, then it will intersect with vicinities of at least N public sites, and the attackers have to tell the real location attribute of message issuer from at least N candidates. The property also guarantees all users in the cloaked region form a k-Anonymous set.

There are two main steps to realize (N,k)-Diverse anonymity: (1). Generate a diverse spatial cloaking region; (2) Render the region to be a k-anonymity region.

Correspondingly, we design an **N-Location diversity Cloaking** Algorithm (**N-LDC**) to realize step1, and then propose **Node Increment Cloaking** (**NIC**) and **Region Increment Cloaking** (**RIC**), which are derived from traditional k-anonymity cloaking, to realize step 2. Further, we improve our N-LDC to **Secure-NLDC**, to defend Central PS attack, which strengthen the robustness of (N,k)-Diverse anonymity.

TABLE 1. LOCATION RELATED INFORMATION OF USER

User ID	Nearest PS	Location attribute
A	Shoe factory	factory worker
B	Shoe factory	factory worker
C	Shoe factory	factory worker
D	Middle School	Student, teacher
E	Middle School	Student, teacher
F	Restaurant	cook, consumer
G	Restaurant	cook, consumer

Table 2.location related information of cloaking region

Cloaking region	contained users	Possible nearest PS of users in the region
R1	A, B, C	Shoe factory
R2	C, D,E,F	Shoe factory, middle school, restaurant

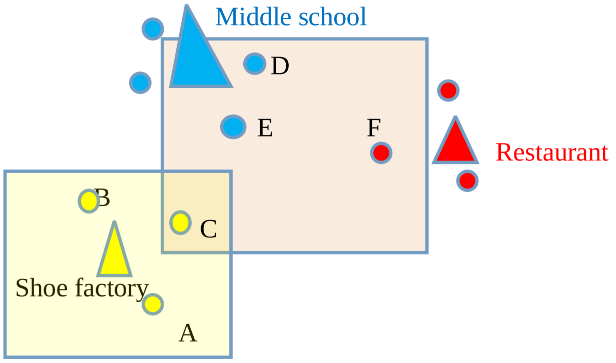

Fig.1 example for table 1

The paper's contributions are as follows:

- We discover a novel privacy issue: *Location Attribute Leakage*, which traditional K-anonymity fails to guide an effective solution.
- We propose a novel but effective protection scheme: **(N,K) diverse anonymity** to defend the location attribute leakage while guarantee the user's query to be k-anonymity.
- We propose a novel but efficient metric and method to evaluate *location diversity* based on voronoi diagram.
- We give algorithms, RIK and NLDC, which can render a desire cloaked region's as minimum area as possible.
- We address two kinds of messages' location attribute privacy: message whose user whose exact location is private; message whose user exact location is public known.
- We implement an experimental anonym based on our techniques, which can offer three security schemes for user to choose based on their need. They are Capser cloaking[26], PN_LDC(plain NLDC), ANLDC(advance NLDC) .
- We conduct extensive theoretical and experimental analysis of our techniques, which suggest that our scheme is highly in effective in guaranteeing the user's privacy in terms of time, space and robust to attacks.

The rest of the paper is organized as follows: Section II presents some basic preliminaries .Some related work is presented in Section III. Our algorithms are proposed in section IV. Then, extensive analysis of security and performance of our scheme are discussed in section V, in which an improved algorithm is designed to better defend privacy attacks. The results of our experiments are illustrated in section VI. Finally, section VII concludes our paper.

II. PRELIMINARIES

Before give the definition of (N,k)-Diverse anonymity, we introduce some definitions.

Definition 1.PS (PS): A PS is a public place where the people who located may have something to do with function of the site.

For PSs that locate in different positions,they are regarded as different PS even if they share some common attribute.

Definition 2.Location attribute of user (LA): A user's location attribute describes the user's location relationship with PSPS he or she is nearest to in a period

For example, in a period T, an individual whose nearest PSPS is hospital will be a patient or doctor. An individual may have different location attributes over different time.

Definition 3.Vicinity of PS (VOP): VOP of a specific PS is a vicinity region around the PS. anyone who located in the VOP, the PS would be his or her nearest PS.

Note that every PS has its own VOP and at every time one people can locate in only one of vicinity of a PS. For example, at time t, a people is in the vicinity of a hospital, then the hospital is the nearest site of the people at time t.

We will offer a metric to compute the VOP for every PS in a certain range of area.

Definition 4.N Location diversity for a region: a region R is said to be N location diversity if R covers parts of N VOPs of N different PSs. The condition's formal description is:

$|\{ PS| \forall PS, R \cap PS.VOP \neq \emptyset \}|=N.$

Definition 5. N Location diversity for a query message: A message Q is said to be N location diversity if the message's potential issuer (the user who may issue the massage) may be from any one of N different PSs' VOP. The formal description is : $|\{ A.LA| \forall A \in$ Users who may issue Q$\}|=N.$

Definition 6. (N,K)-DIVERSE ANONYMITYK-Anonymity: A query message is said to be (N,k)-Diverse anonymity if and only if itsatisfiesthe following two conditions: (1). k-Anonymous: it can be related to specific user not higher than 1/k. (2). N Location diversity: it must satisfy N Location diversity defined in *Definition 5*.

Note that the '*N diversity*' differs from the '*k-Anonymity*' in that it does not guarantee the message being related to a specific target (location, people, etc.) with probability of 1/N. However the anonymity focus on undistinguishable and a specific probability is guaranteed. However our algorithm *secure NLDC* can render the message being related to a specific target around probability 1/N, we will give the details in the section V.B.

Definition 7.Region R: A Region R is a rectangle in two dimensional Euclidian space, which is defined by two endpoints S and E of its major diagonal:

$$R=(S, E)$$

Where S=(x1,y1), E=(x2,y2).

A. *Metrics*

To achieve N location diversity for a group of users, we need to gather users from at least N different places (N different PSs) by classifying the nearest PSs. Before, the PSs' VOPs of all specific users belongs to should be determined. We adopt the clustering method based on Voronoi Diagram[19]. We first divided the space into subregions using the Voronoi rule, the Voronoi seeds are a set of the given PSs. The Voronoi seeds and the Voronoi cell are denoted by *Vseed* and *Vcell* respectively. According to the property of Voronoi diagram, we have the following lemma.

Lemma 1: *Vcell is Vop of its Vceed.*

Proof: According to the property of Voronoi diagram, given a Voronoi diagram of a set of Vceeds and a query point P, P's nearest *Vceed* is the *Vceed* whose Vcell contain P. In other word the *Vcell* is VOP of its Vceed, so that a certain *Vceed* is the nearest neighbor to all query points which fall into its Vcell.

According to Lemma1, given a Vornoi diagram of a set of PSs, users' location attribute can be classified by the *Vcells* they belong to. Thus the location diversity can be measured by the following two concrete metrics. They are respectively based on two type of user: (1) the user whose exact location is public known; (2) the user whose exact location is private.

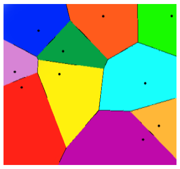

Fig2. Voronoi diagram

Metric1. Metric for (N,k)-DA of the users whose exact locations are private. In order to guarantee the query message of user being (N,k)-DA, the cloaking region must satisfies two conditions:

(1). The users in the region form a k-Anonymous set[7,17,27.];

(2).The region intersects with parts of at least N *Vcells* of N PSs: |{ PS| ∀ PS; R ⋂ PS.*Vcell* ≠ ∅ }|=N.

Lemma2: *When user's exact location is private, the R that satisfies the Metric1 can guarantee* (N,k)-DA *for user in it.*

Proof. Because the user in R is k-Anonymous, then a message can't be related to any one of them higher 1/|{user| ∀ user located in R}|. Plus at least K users' location is private, so a message from the region can be from anywhere of the region because of the condition (2), so a message from the R satisfies(N,k)-DA.

Metric 2. Metric for (N,k)-DAof users whose locations are public known. In this scenario, in order to ensure a query message of user being (N,k)-DA, there are two conditions must be satisfied by a corresponding cloaked region R:

(1). The users in the region form a k-Anonymous set[7,17,27];

(2). These k(or more)users must from N *Vcells* of N different PSs: |{ User.nearest PS| ∀ user located in R}|=N.

Note that the different between conditions (2)in Metric 1 and Metric 2 is that Metric1 does not require every parts of N cells must contains a user.

B. *Architecture*

We adopt User-Broker-Server architecture (see Figure3) to take a further analysis of (N, k)-DA. K-Anonymity model have been extensively studied in [7,17,27,29]. The anonym in (N, k)-DA is in charge of the following tasks:

(1) Cluster users by the PSs they are nearest to.

(2) Form an anonymous region for user which make the message satisfies (N,k)-DA according to the user's privacy requirement.

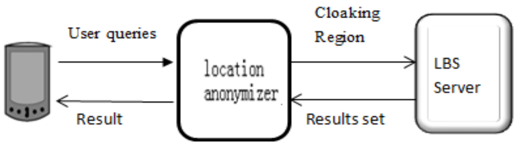

FIGURE. 3. ARCHITECTURE OF TRADITIONAL LBS SYSTEM

The first thing is to store Voronoi diagram for all the PSs in a given area in Location anonym. We denote the Voronoi diagrams of PSs in region A as *Vmap*(A)**.** *Vmap* is Graph structure, which is described by *Vmap*(A)={*Vcell*(PSi) | ∀ PSi fall in region A}.

Every PS is represented as a 2-dimension Point<x, y> for simplicity, and *Vcell(PS)* is a polygon that bound the PS.VOP using PS as *Vseed*. *Vcell* can be represented as a graph, Vcell.E represents the edges of *Vcell*, *Vcell*.v denotes the vertex of *Vcell*.

Besides *Vmap(A)*, we need an index of all *Vcells* in anonym. We index the *Vcells* using R-tree [19], the leaf of the R-tree is MBR of every *Vcells* of *Vmap*. We term the R-tree as *RVmap(A)*.

In terms of efficiency, there is no global anonym for LBS user. Anonym is in charge of a certain range of region, if a user enters this region he or she can register his or her basic information in anonym by submitting a *user profile* to anonym through secure links. User is 6-element tuples, which can be denoted as *User*<ID, nearest PS, N, k, location, CoverRatio>, where the N, K is the parameter for (N, k)-DA, Cover ratio is a security level parameter we would discuss it in latter section. Note that the N, K, coverRatio is a secret share between user and anonym(broker).

III. RELATED WORK

A. *Data structure for k-Anonymity cloaking algorithm*

Quadtree[26] structure is the main data structure for describing the interval cloaking and the Casper cloaking. It recursively partition the space into smaller quadrants until each quadrant contains points less than a predefined threshold. A quadrant which is undividable and contain points less than predefined threshold is leaf node of Quadtree. We use Figure 4 to illustrate the working principle of Quadtree.

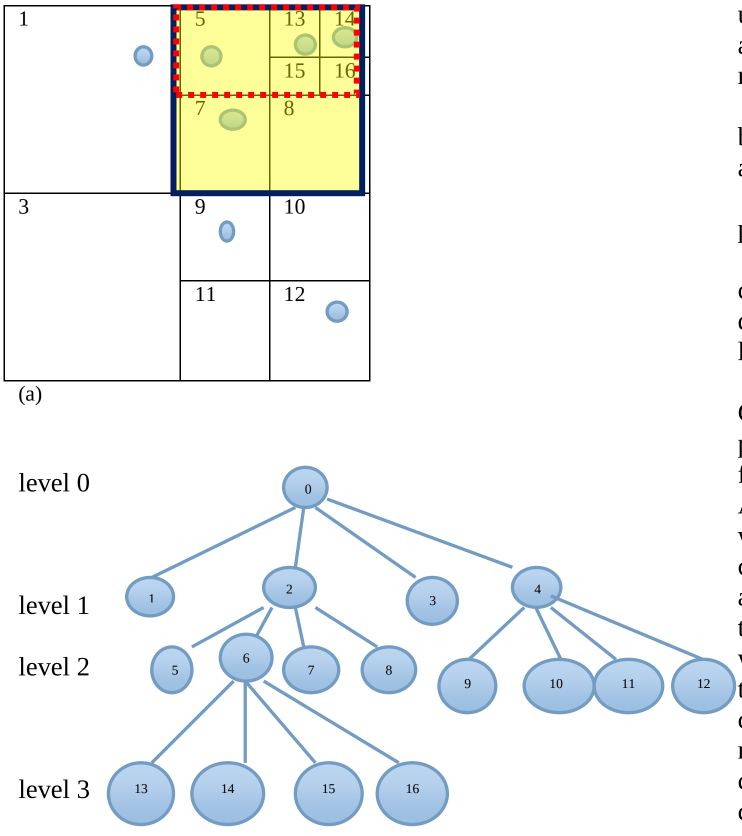

FIG .4.: EXAMPLE OFQUADTREEPARTITION

Figure 4(a) demonstrates a Quadtree partition where the threshold is 1 point/ quadrant. The whole region have seven points, so the area is firstly partitioned into 4 quadrants, which are in first level of Quadtree and showed in (b) at level1. Then the quadrant2 and quadrant4 have more than one point so it continue to be divided, where the quadrant2 divide into quadrant5, quadrant6, quadrant7, quadrant8 and quadrant4 is divided into quadrant9, quadrant10, quadrant11, quadrant12. These new quadrants are showed in Figure 4(b)'s level 2. However, because quadrant 6 contains more than one point, it continues to be divided, generating new quadrants in level 3.

B. *k-Anonymity cloaking*

Interval cloaking[27]: interval cloaking is first k-Anonymity cloaking algorithm. The anonym stores a Quadtreee of users in a given area. Given a user point, the interval cloaking algorithm traverse the Quadtree in up-to-down way until it first meets a quadrant that contains less than k-users but contains the user point. Then the algorithm selects the quadrant's parent as cloaking region. In Figure 4, for example, we assume the query point is in quadrant 13, and k is three. The algorithm first exams the top level of Quadtree finding that there are seven users, which is larger than k, so it traverses down to the level 2 to find the quadrants in level 2 that contain the query point. The quadrant 2 contains the query point and is in level 2, so algorithm exams it, finding that it also contains more than k users, so it continue traverse down, until it traverse to the quadrant 6 which contains less than k users. It finally selects the quadrant 6's parent node quadrant 2 as cloaking area. The final cloaking region is the yellow rectangle with black frame showed in the Figure 4(a).

Casper cloaking [26]: Casper cloaking is also Quadtree based algorithm, it differs from the Casper cloaking in two aspects:

(1)It traverses the Quadtree in a bottom-up way. The start point is the lowest quadrant node that contains query point.-

(2)It exams whether the current quadrant's neighbor can combine with the quadrant to include k or more users before directly choose the parent as cloaking area, hence it render less cloaking area.

Now, we run the same example as above to illustrate the Casper cloaking. The smallest quadrant that contains query point is quadrant 13, so the algorithm start from quadrant 13 finding that the quadrant contains less than k (k=3) uses. The Algorithm hence exams whether the quadrant 13 can combine with its neighbor quadrant to reach k users, if it can, the final cloaking region would be a region just cover the quadrant 13 and the neighbor, otherwise the algorithms traverse upward to the parent node. In this example, quadrant 13 can't combine with any of its neighbor to reach the goal, so the algorithm traverse to the parent node: quadrant 6. The quadrant6 contains less than k users but it can combines with its neighbor quadrant5 to reach k users. So the algorithm combines the quadrant 5 with quadrant6 to form a final cloaking region, which is showed as yellow region with red dotted frame in the Figure 4(a).

All the work mentioned above didn't consider PS reference attack, computing user's location attribute by referent PS he is located in or nearest to. For example, a user who locates in a hospital has a higher probability of being a patient. When a cloaking region is related with a specific PS, then all the previous work fails to protect the users location attribute from being exposed. To conquer this implication, we propose a new privacy property, (N,K) DA, instead of spatial k-anonymity to guarantee user's privacy. Algorithms in following sections are proposed to realize this novel but stronger property.

IV. ALGORITHM

A. *Data structure*

Our goal is to meet (N,k)-Diverse anonymity for LBS users while make the generated ASR（ anonymous spatial region） small as possible. There are two ways to be chosen.

(1) Generate a k-Anonymity region for query user first, and then if the region does not satisfy N location diversity, render it to satisfy N location diversity.

(2) Generate a region that satisfies N location diversity for user, if it does not satisfy k-Anonymity then render it to satisfy k-Anonymity.

The second way is developed in this article and the first method is to be completed in our future work.

All of our algorithms are based on two structures: Quadtree (of users) and Voronoi diagram(of PSs). Quadtree is used to render a region to satisfy k-Anonymity, whileVornoi diagram is used to render a region to satisfy N location diversity. The main function of our algorithm is illustrated by pseudo code in Figure 5.

```
(N,K)-D A _main(QDT node, Vmap v, User u)
1.Form a N location diversityregion R for u;
2.if R satisfiesK-Anonymity then
3.    return R;
4.else then
5.    render R to satisfies K-Anonymity
6.    return R.
```

Fig. 5 Main function of (N,K)-DA cloaking

In the algorithm of main function of (N,k)-DA cloaking, the parameter QDT denotes the structure of Quadtree, node is the root of Quadtree of LBS users, Vmap denotes the Voronoi diagrams structure, v is the Vornoi diagrams of PSs in a given region. The parameter u denotes the query user. (N,k)-DA_main demonstrates the procedure of using method2 to generate (N,k)-DA cloaking region for query user. The main function calls two sub functions; they are *k-Anonymity cloaking* and *N location diversify cloaking* respectively. We discuss these two cloaking respectively in following sections.

B. *New k-Anonymity cloaking algorithm*

Two kinds of k-Anonymity cloaking technology, namely,*NB(node based) cloaking algorithm* and *RB(region based) cloaking algorithm* are introduced. For *NB* cloaking algorithm, both the processing unit and the input are quadrant, hence both the Casper cloaking and the Interval Cloaking are node based cloaking algorithm , while the *RB* cloaking algorithm is to render an input region, which is arbitrarily distributed and may not overlaps with any quadrants, to satisfy k-Anonymity property..

1) NB cloaking algorithm

For this algorithm, given a query point u, we first find a quadtree node that contains query point as an initial cloaking region for u, if it contains less than K uses, we will enlarge it based on certain rule to satisfy k-Anonymity.

The Interval cloaking [27] and the Casper cloaking [26] are both *NB* algorithms. Bycombining the features of them, a new *NB* cloaking algorithm is proposed, which is described in Table 6. The proposed algorithm can both quickly find the cloaking region and generate a small cloaking area as possible. Because it works in a **node increment** way, we name it **NI** cloaking algorithm.

In NI cloaking algorithm, it traverses in top to bottom way to find the quadrant that first appear to have less than k users, instead of using its parent as cloaking region we exam whether it can combines with one of its neighbors to form a k-Anonymity region. Firstly, NI cloaking algorithm judges if the input node is null, if it is null, then find the lowest Quadtree node that contains user(line 1~2). After that if the querynode contains more than K users, the algorithm will return the region pointed by the node (denoted as node.R) as cloaking region(line4~5), else if the region pointed by the root node of the tree contains more than K users, the algorithm traverse the tree in the same way with Interval cloaking until it encounter the first node that contains less than K users(line 8). The nodei in the algorithm denotes the first node that showed to be less than K users to the algorithm. The algorithm then starts from the nodei using the same traversing strategy as in Casper to traverse the tree to generate the desire cloaking region (line 9~11). If the region pointed by root node contains less than k users, the algorithm return the region(line13~14) .

```
NI( Qdtrootnode, intk,Qdtquerynode)
1. ifquerynode=null then
2. querynode = lowest node of quadtree that contains user
3. end if
4. ifquerynode.num>k then
5.     Return querynode.R;
6. end if
7.ifrootnode.num>k then
8.     Traverse the tree in the same way as the interval until
find first node nodei that contains less than K users but
contains query point.
9.     if nodei can combines with anyone of its neighbor to
contain K or more users  then
10.      Combine with the one of its neighbor which can offset
with it to reach K or more contained users.
11.      else
             choose the nodei.parent.R as cloaking region
12.    end if
13. else
14.     return rootnode.R
15.  end if
```

Fig. 6.Node based cloaking algorithm

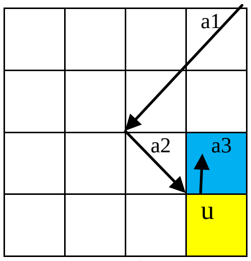

Fig.7. Example of NodeIncrek

Figure 7shows an example of our algorithm, the Quadtree in the Figure 7 has totally 3 levels. Its root node points to the whole area of the region, arrow a1, a2, a3 show the algorithm's traversing trajectory. At first the whole region contains more than K users, so the algorithm traverses down(a1 shows) to the sub region(contains user) pointed by one of the rootnode's son, however the sub region contains more than K users, so the algorithm continue to traverse down(a2 shows) to a region pointed by a grandson of root node which contains less than k users but contains query user(the node is yellow node in the Figure), so the algorithm stop traversing down to process the node in the same way as in Casper to form an k-Anonymity area (blue area + yellow area).

2) *RB cloaking algorithm*

All previous k-anonymity cloaking are NB cloaking, the RB k-anonymity cloaking is proposed for our new privacy property.

a) Motivation of developing the algorithm

Firstly (N,K)-DA generates a region that satisfies N location diversity property, then call k-Anonymity cloaking function to render the initial region to satisfy k-Anonymity property if it doesn't satisfy it. So in the (N,K)-DA, we need to form a k-Anonymity region based on an input region that is arbitrarily distributed, which may not be pointed by any quad tree node.

b) Algorithm description

Inputting an arbitrary region, the region based cloaking algorithm aims to render the region to satisfies k-Anonymity, where k is the user customized parameter. The algorithm is described in Figure.8.

Firstly, we introduce our *merge* rule, which is frequently used in the RI algorithm. The rule is described in Figgure8 , named MR cloaking algorithm, short for Merge Region. The rule selects minimal x1,y1 from two region's endpoints to form a new left bottom endpoint S, and select max x2,y2 from two regions endpoints to form a new top right endpoint E, the two new endpoints define a new Region(see definition 7).In the Algorithm E(R.v1, R.v2) is the major diagonal line of the new region.

```
MergeR(region R1 , Region R2)
    minx=Min(R1.x1, R2.x1);
    miny=Min(R1.y1, R2.y1);
    maxx=Max(R1.x2, R2.E.x2);
    maxy =Max(R1.E.y2, R2.E.y2);
    R.v1=(minx, miny);
    R.v2=(maxx, maxy);
    return R;
```

Fig. 8. Merge region rule

Figure 9gives an example.

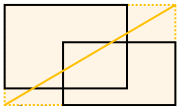

Fig. 9. Example of merge

Definition 8.MinCover Quadrant(MinCover): Given a region R and a Quadtree with all users. Mincover refers to a certain node of the Quadtree which can cover the whole area of R and whose children quadrants (if it has) can't fully cover R.

Figure 10gives an example for Definition 8: region R1 intersects with Qudtree and its minimal cover quadrant is the yellow quadrant of Quadtree, which is in the second level of the Quadtree.

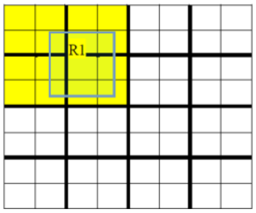

Fig. 10. Example of Definition 8

```
RI(Qdtrootnode, QdtMinCover, int K, int threshold, Region IniR)
if mincover.num<k then
    R=NodeIncrek( rootnode, k, MinCover1);
Return R;
else
     for i=1 ~ 4 do
IniR[i]=IniR ∩ MinCover1.children[i];
MinCR[i]= MinCover of IniR[i]
     End for
Num=0;
       While Num<k do
Num=0;
             For i=1 ~ 4 do
If MinCR[i]< MinCover1.children[i] then
CRnum[i]=MinCR[i].Num;
MinCR[i]=MinCR[i].parent;
                   Else
Num=CRnum[i]+Num;
              End for
          End while
          R=∅;
```

```
21.     For i=1 ~4 do
22.         R=MergeR(R, MinCR[i]);
23.     End for
24.     Return R;
```

Fig. 11.Algorithm of Region increment cloaking

The goal of the region based cover is to render the input region to satisfy k-Anonymity. The algorithm work accord to following steps:

(1)It computes the MinCover for the Initial region, which is denoted as MinCover1 in the algorithm.

(2)Then, if number of users in MinCover1 less than K, it will use the Mincover as input triggering Node based algorithm to form K-Anonymity region (Line 1~3).

(3) Else, it computes a k-Anonymity region contained in MinCover(in a increment way)and merges this region with initial region to form a new region which definitely satisfies k-Anonymity. (Line 4 ~ 25)

The idea of (3) is that for every child of MinCover1Mincover for every part of initial region is computed. We termed these MinCovers as SubMincover of initial Region. In the algorithm, IniR[i] denotes the part of Mincover which fall into one of the Mincover's children. In Figure 12, for example, the whole rectangle is the MinCover of Blue Region. The part of the blue region in the ith Mincover's child (denoted as MinCover1.children[i]) is named IniR[i]. The min1, min2, min3, min4 are SubMincovers of the blue region and are MinCover of IniR[i] respectively, where i= 1,2,3,4.

After computing SubMincovers for initial inputed region, we exam whether the total number of users in these SubMincovers exceeds K or equal to K. If yes, we merger all these Submincovers with initial region to form final cloaking region (Line 21~24).Else we expand all the SubMincovers recursively until total users they contain reach k (Line 10 ~ 18). In every loop, we expand all the SubMincovers to their parent nodes respectively, if expended areas involve k or more users we merge all these expanded areas to get final cloaking area. In Figure 9, if total number of users in min1, min2, min3, min4 reaches k, we merge these 4 areas using rules in algorithm MergeR. Else we expand min1, min2, min3, min4 to their parents node respectively, exam number of contained users again, if it reaches K, merger them together, else continue expanding them in the same way. Note that for each SubMincover, it can't be expanded to surpass whole area of MinCover1's children in which they are located (In the algorithm, MinCover's four children are denoted as Mincover1.child[i], where i=1,2,3,4.).

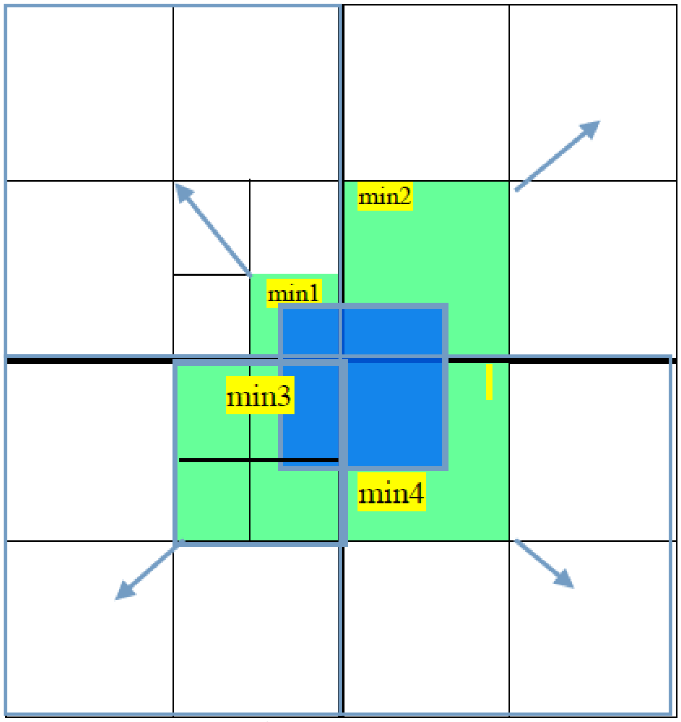

Fig. 12.:Example of Region increment cloaking

C. *N Location Diversify Cloaking Algorithm (N-LDCA)*

In this section we introduce an algorithm termed N-LDCA, Which can generate a region R satisfying N location diversity while its cloaking area is minimal as possible. If the R doesn't satisfy the k-Anonymity property, it will become the input region of function *RI*, which will render the R to satisfy the k-Anonymity.

In order to form an *N Location Diversity(N-LD)* region, the necessary steps are:

1. Find the set of N PSs whose Vcells are as concentrated as possible
2. Generate a region R that intersects with all Vcells of these PSs and R is minimal as possible.

Now we give a necessary definition that provides a criteria for finding candidates PSs in step 1.

Definition 9.Minimal Border Distance(MBD): MBD between a Vcell(p)of PS p and query point u is denoted as MBD(u,Vcell(p)). A circle centric at u, with MBD(u,Vcell(p)) as radius can just tangent with VOP of p.

For easier understanding, we illustrate the Definition 9 in Figure 13.

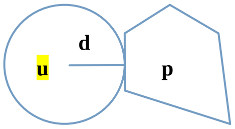

Figure13.Example of definition 9

In the Figure 13, u is query point and is also the centric of the circle, p is a PS, the polygon that encompasses the p is p's VOP, the circle centered at u is tangent with the polygon. The circle's radius d, as the Definition 9, is the MBD(u,Vcell(p)).

However, to compute the exact MBD is quite complex, we hence give an approximate algorithm to computer MBD. We borrow a metric from[28] called Minimal Distance(MINIDIST) to compute the approximate MBD.

There we cite the MINIDIST [28] and restate it in Definition 10.

Definition 10. Minimal Distance: the MBD between a query point u and a region R is denoted as MINDIST(u,R), which is calculated as:

$$\text{MINIDST(u,R)=} \sum_{i=1}^{n} | ui - ri |;$$

Where ri={
R.Si if ui<R.Si;
R.Ei if ui>R.Ei;
Ui otherwise
}.

In two dimension case, the n in Definition 10 is 2, and R.S1=R.x1, R.S2=R.y1, R.E1=R.x2, R.E2=R.y2.

Ana ïve idea to find N candidates PSs is to find N PSs who have N nearest MBD to query point, however calculating the MBD is very complex. Because MINIDIST(u, R) is computational efficient, we find out the N PSs by the MINIDIST. The method is that we pre-compute the MBR for all PSs in a given region and store it in database, when we need to find N PSs, we select them according to the MINIDIST of these PSs to query point u, in other word, the final selected N PSs are those whose Vcell's MBR has N smallest MINIDISTs to u.

In order to form NLDR, our method is to generate a circle region that intersects with N Vcells firstly, and then form the MBR of the circle. So the circle determines the final NLDR. The following three points should be concerned in NLDC.

Firstly, the circle's radius is at least larger than the distance between centre point and the Nth MBD Vcell to the centre point. This condition guarantees that the circle can at least intersect with N vcells. Secondly, as we know, the distance between a point and its Nth MBD Vcells is computational complex and not necessary, we hence, use the approximate MBD to substitute the exact MBD.

Finally, we are not going to pursue utter optimization on the anonym side, for its limited resource comparing with the LBS servers. Little optimization in many cases contributes so trivially on the query process on LBS server side that we can ignore the contribution of utter optimization.

Based on the above guidelines, we compute radius for the circle based on Approximate MBD. We now first give a necessary definitionfor our approximate MBD.

Definition 11. Corresponding Edge(CE): Given a query point P, and two dimension region R, the CE((x1,y1),(x2,y2)) of R is calculated as the following algorithm depict in Figure 14:

```
ComputeCE(Region R, Edge E, point P)
1.If p is in R  then
2.          CE= any one edge of R's edge;
3.      Return CE; 4. end if
5.   For i=1~2 do
6.      If p.[i]<R.S.[i] then
7.          CE.[1][i]=R.S.[i];
8.          CE.[2][i]=R.S[i];
9.      Else if  p.[i]>R.E[i] then
10.         CE.[1][i]=R.E[i];
11.         CE.[2][i]= R.E[i];
12.       Else then
13.           CE.[1][i]=R.S[i];
14.           CE.[2][i]= R.E[i];
15.   end if
16. If  CE[1][1]=CE[2][1]&&CE[1][2]=CE[2][2] then
17.      CE= one of R two adjacent Edges whose common endpoint is (CE[1][1],CE[1][2]);
18     Return CE;18. Else
19.     Return CE;
20.  end if
```

Fig. 14. Compute corresponding R.

The algorithm in Figure 14 is to choose one edge of R as CE. In the algorithm p.[i] denote the pi, for example if P=(x, y), then P[1]=x, P[2]=y.

The algorithm chooses one edge of Recording to location relationship between query point and region R. if query point is in R then it will choose one of any edge of R(see line 1~3) and Figure 15 (a) show the case, the edges in red are the candidates CE for R. if query point outside the R, and P.[i]> R.S[i]&&P[i]<R.S[i] where i=1 or 2, just as Figure 13 (b) shows, in this case there are two parallel edges whose extents (dotted line in (b)) contain p, and the CE is an edge that is perpendicular with these two parallel edges and nearest to P. if there are no parallel edges of R whose extents contain p, and p is outside of R, the Figure 15(c) depict the case. In this case, two candidate CEs, which is calculated by line 5~14, intersects with each other at a point, this point is nearest vertices point of R to p.

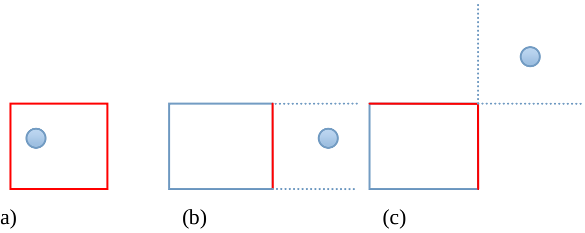
(a) (b) (c)

Fig.15. Example of choosing CE

In fact when compute the MINIDIST we have gain the location relationship between query point and some regions, with this information we can soon find out the CE.

We assume that some vertices of Vcell which is on the edge of MBR of the Vcell are prerecorded, in other word, the structure of Vcell MBR's edges contain the which vertices of Vcell on them. Note that the all Vcells are convex polygons, so there are at most two vertices of one Vcell on one edge of its MBR.

Now we give the Definition of approximate MBD.

Definition 12.Approximate MBD(AMBD): given a query point p and MBR R of a Vcell v, the AMBD is the distance between the p and p’s nearest vertices of v which is(are) on the CE of R.

We use an example in Figure 16to illustrate the definition 12.

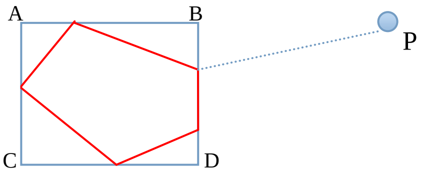

Fig. 16.Example of AMBD

In Figure 16, rectangle region $\overline{ABCD}$ is MBR of a Vcell(red edge polygon), $\overline{BD}$ is CE of the region, there two Vcell vertices on the $\overline{BD}$, the dotted line link one that nearer to p with p. The length of the dotted line segment is AMBD.

Lemma 3.*Given a Vcell v, a query point p and AMBD(v,p) d, the MBR of circle which center at p with radius d intersects with v.*

Proof: In definition 9, the circle center at p with radius MBD(v, p) tangent with v. Because AMBD>= MBD, so the circle in Lemma3 at least tangent with v, and its MBR intersects with v.

Now, we give our first N-LDCA

1. Select N Vcells from a given Vmap who have N smallest MINIDIST to query point P.
2. Sort the N Vcells by its AMBD to query point in increasing order.
3. Record Nth Vcell’s AMBD to P, we denote it as ND.
4. Generate a circle which is centered at P with radius ND.
5. Return the MBR of circle in step 4 as NLDR of query point.

According to Lemma 3, the cloaking area in above algorithm can at least intersect with N Vcells, it applies to NLDR too. However the region may intersects triviallywith some Vcells, to overcome this shortage we can extend it by a user customized value: *coveratio*, which is one of parameters of user profile.

Figure 17shows our algorithm for the expansion. Assume the polygon in red edge is the Vcell that have Nth AMBD to u. uis the center of the blue circle. The solid line segment that link u with one of vertices of Vcell is Nth AMBD to u, hence the blue circle is intersects with at least N Vcells. The dotted line is extension of the Nth AMBD, this extension is customized by user. The orange rectangle is MBR of the circle with extendedradius, the MBR is also the final NLD cloaking area of our algorithm.

If NLDC does not satisfy K-Anonymity, we use the region based cloaking algorithm, which is introduced in previous section, to render it to satisfy K-Anonymity so that the final cloaking region is (N,K)-DIVERSE ANONYMITY.

However the NLDR which is MBR of circle centered at query point is not a secure cloaking region for user, because it would suffer from *Center PS attack*. We would give detail descriptions and a counter measurement to the attack in next section.

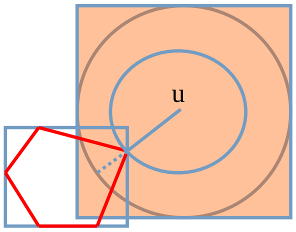

Fig. 17.Simple Example of algorithm NLDC.

V. PRIVACY ANALYSIS

A. *Center PS Attack*

In section III we first restate our algorithm for forming a (N,K)-DA cloaking region for query user, there are two main steps:

(1) Form N-LD cloaking region for user.
(2) If N-LD cloaking region doesn’t satisfy k-Anonymity, we then render it to satisfy the property.

However, the N-LDR generated by N-LDCA algorithm we described in section III.C is actually the MBR of a circle, which is centered at query point, in other word, the NLDR’s center is user point so that the attack can deduce the user’s real location.

Even though the final cloaking region may be outcome of k-Anonymity cloaking function using the NLDR as input, rather than original NLDR, the center of final cloaking region may have high probability of being near or in the Vcell that user is in, especially when the k-Anonymity function need only to expand little from the original input region to include K anonymous users.

We call method that privacy attacker use the center of cloaking region to deduce the location information about query user as *Center PS Attack*. Obviously, the (N,k)-DA region generated by our previous algorithm is vulnerable to *Center PS Attack (CPA).*

Figure 18illustrates an example of *Center PS Attack*. In the Figure , the center of cloaking region is in yellow Vcell, which has high probability of being the query user’s LA.

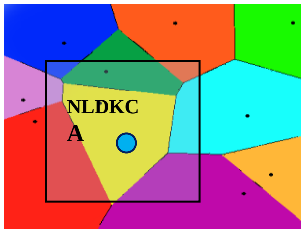

Fig. 18.: Example of Center PS Attack

B. *Counter measure to CPA*

So far, we know that the NLDR generated by NLDC algorithm in section III.C is vulnerable to *CPA*, because the NLDR is centered at real query user’s location. One intuitive method, which also renders less NLDC modification, to defend *CPA* is to substitute real query point with a fake point.

If we can make the attacker who uses the *CPA* to deduce the Vcells in which user is located with a probability being not above 1/N, we actually defend *CPA*. Based on the principle, we propose a new NLDC, called Secure NLDC.

```
Secure NLDC(point P,Vmap VM)
1. Select N Vcellsfrom a given Vmap who have N smallest MINIDIST to query point P;
2.Randomly select one vcellvfrom N vcells in line1;
3. randomly select a user in the vas fake query point P1;
4. N=P.N
5. NLDR=NLDC(P1,N,VM);
6. if radius of NLDR's inscribed circle <dist(p,p1) then
7.     Generate a circle same center with the NLDR but with radius of dist(p,p1);
8. expanded the circle with P.coveratio;
9.   form a MRB R1 of the expanded circle in line 8
10.         return R1
11. else
11.   return NLDR
12. end if
```

Fig. 19: Secure NLDC algorithm

In the algorithm we first select N Vcells that have N smallest MINIDIST to query point P, and then randomly select a Vcell of these N Vcells, choosing one user point in the Vcell as fake point. Finally, run the NLDC in section III using the fake point as input, but the parameter N is the original user P's N. In order to guarantee the final NLDR cover real query user, the circle of final cloaking area must with radius larger than dist(p,p1) (Line 6~8), in the Line 8 we extend the circle with P's coveratio is to guarantee further security.

C. *Security analysis of our technique*

Assumptions:

(1) Anonym is secure to LBS users, it has secure links with LBS users. LBS server is public and not secure to its user.

(2) K value, N value and Coveratio of (N,K)-DA are customized by users, and they are secrets shared only by user and anonym.

(3) Every user's ID is removed by anonym and replaced with pseudo ID.

Theorem 1.*Given a N-LDR that is generated by Secure N-LDCA, the origin PS of query message can't be assured by CPA with probability higher than 1/N.*

Proof. Because the *CPA* attacker doesn't know the N and K, so he can only estimate N by the number of PS that cloaking region intersects with, and we know this number N1>=N. According to *Secure NLDC* algorithm, fake PS is chosen from one of query PS's N smallest MINIDIST neighbors randomly, for the same reason, the query PS is one of N1smallest MINIDIST neighbors of fake PS. So, every PS with which the cloaking region intersects would be the potential PS to attacker, with same probability, so the success rate of *CPA*is about 1/N1<1/N. The k-Anonymity cloaking may even further contribute to forming the final cloaking area, which makes the success rate of CPA further less.

D. *Computational analysis of our technique*

The main computation overhead of NLD cloaking is to find N Vcells whose MBR are N smallest MINIDIST to query point, [28] gives the computational bounding of find N minidist MBR as O(nlog n), n is total number of PSs in a given region. The computational overhead of our region based k-Anonymity cloaking is at most four times of interval cloaking[27], because when MinCover have more than K-users, we would find a k-cloaking region inside MinCover, the process is equal to run four interval cloaking functions on four children of MinCover respectively. However when MinCover contains less then k-anonymous users, (N,K)-DA_main function would call node based k-Anonymity cloaking function using MinCover as input node, in this case the K-cloaking computational overhead is approach to the traditional k-Anonymity cloaking[26,27,17]. So compare with traditional K-Anonymity cloaking, we need O(nlogn) more time to guarantee our privacy property: (N,K)-DIVERSE ANONYMITY, where n is the total number of PSs in the area that anonym is in charge of.

VI. EXPERIMENT

We have implemented an experimental Anonym based on the technology presented in previous section. Inputs of the Anonym is a 4-tuples< M, Q, U, T>, where M denotes a map which records PSs' distribution and coordinates, Q denotes a Quadtree which records all users' distribution, U is inquirer's profile which contains his/her location and security parameter information, T denotes what kind of Cloaking scheme the user want to use, there are three candidates scheme for user: Casper, PNLDK(Plain PNLDK), ANLDK (Advanced). Note that both PNLDK and ANLDK is secure NLDC, they are differ in that former omits the step(3) in RegIncrek by directly return MinCover as final cloak region as long as it satisfy K-anonymity. So, the PNLDK tend to be more efficient but lager cloaking region than ANLDK.

A. *Experimental result*

In our implementation, we input a region (700*600 m*m), with 30 PSs and 3000 users' distribution. All users and PSs are randomly generated. The visual distribution of our input is showed in Figure 20, where the bigger multicolor points denote PSs, and small points are users.

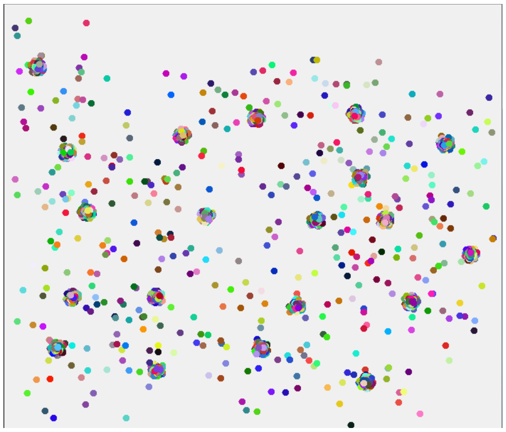

Fig.20 Users and PSs distribution

For better understanding, we use some visual results of our experiment to illustrate the work flow of our scheme.

1) Partition the space by Voronoi rule using PSs as seed. Then form MBR for every vcell, then sorting them according to their MINIDIST to massage issuer.

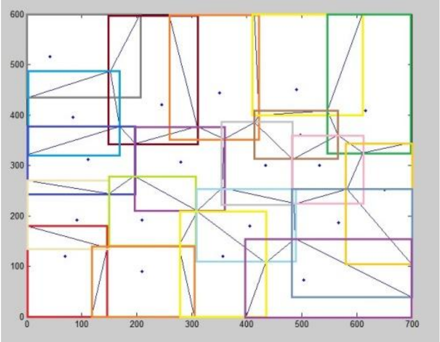

Fig.21. Form MBR for every vcell

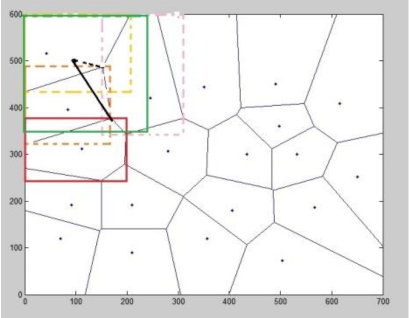

Fig.22. sorting them according to their MINIDIST to issuer

Note that, in the example, our N=4, the black dark line denote the 4th AMBD to the issuer. And the region with green frame is the NLD cloaking region(NLDR) generated by NLDC (because of limited page, we omit the results of Secure-NLDC, which don't effect out experimental correctness).

2) Using our K-anonymity technique to render the NLDR to satisfy spatial k-anonymity.

(1) Partition the space by quadtree method accord to users' distribution.

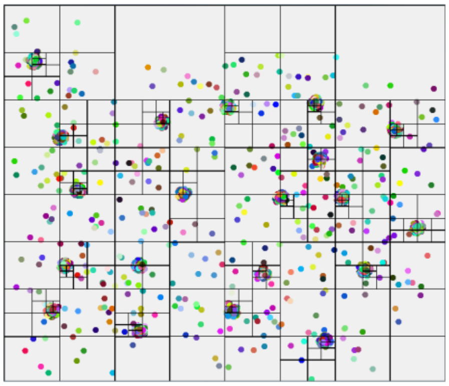

Fig23. Partition the space by quadtree

(2) Using Capser, PNLDK ANLDK to form three different NLDK regions respectively.

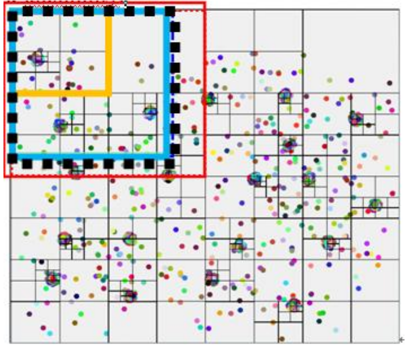

Fig. 24 Three different versions of NLDK region

Note that the region with orange frame is outcome of Casper, the region with black dotted frame is outcome of ANLDK, and the Red frame is outcome of PNLDK. The blue frame region is input NLDR.

B. *Analysis*

We compare the output region in terms of their size over different technology, Casper, PNLDK , ANLDK.

Figure 25 shows the comparison of k-anonymity region size over different K by three technologies, while the figure 26 shows the comparison of NLDK size over various N with a specific K value by the technologies. They both shows that our ANLDK's performance is close to casper, but better than PNLDk. However our ANLDK is more secure than Casper, because the formal can return NLDK cloaking region, the later only return a K-anonymity region.

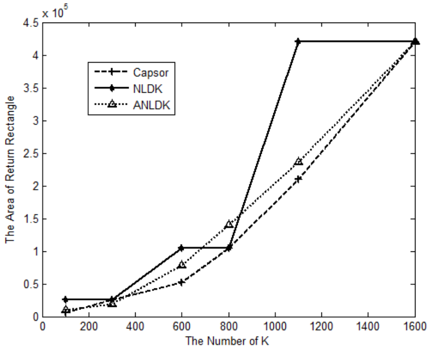

Fig. 25. k-anonymity region size over different K

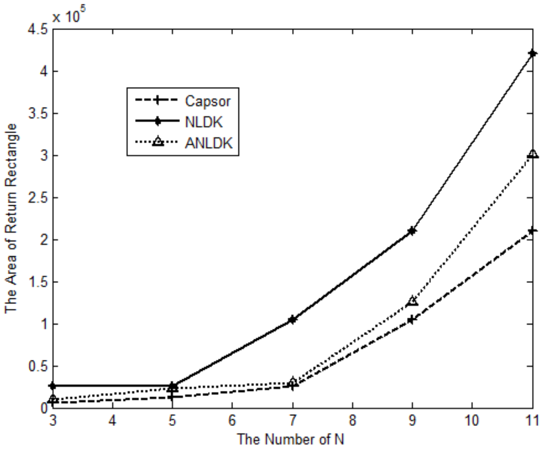

Fig. 26. NLDK size over various N with a specific K

VII. CONCLUSION

In this paper we propose a novel privacy attack: PS reference attack, which the locations attribute of user can be serve as quasi attribute to identify a message issuer. We hence introduce a novel privacy property- (N, K) diverse anonymity to guide an effective privacy protection for users. We then propose some novel algorithms, NIC, RIC, Secure-NLDC, to realize the property. An extensive security analysis and our experimental results show that our techniques are highly effective in protecting user's privacy in LBS system.

REFERENCES

[1] T. XU and Y. CAI, Exploring historical location data for anonymity preservation in location-based services, In Proc of IEEE INFOCOM, 2008.

[2] D. Lin, C.S. Jensen, R. Zhang, L. Xiao, and J.A. Lu, Moving Object Index for Efficient Query. Processing with Peer-Wise Location Privacy, In Proc of Int. Conf. on Very Large Data Bases Endowment, 2011, pp.37-48.

[3] C.-Y. Chow, M. F. Mokbel, and X. Liu, A Peer-to-Peer Spatial Cloaking Algorithm for Anonymous Location-based Services, In ACM International Symposium on Advances in Geographic Information Systems, 2006.

[4] W.E. Winkler, Advanced Methods for Record Linking, Section on Survey Research Methods (American Statistical Association), 1994.

[5] N. Koudas,B. C.Ooi,K.-L.Tan, and R.Zhang,Approximate NN queries on streams with guaranteederror/performance bounds, In Proceedings of the InternationalConference on Very Large Data Bases, 2004, pp. 804-815.

[6] T.XuandY.Cai, Feeling-based location privacyprotection for location-based services, In CCS'09:Proceedings of the 16th ACM conference on Computer andcommunications security, ACM, New York, NY, 2009, pp. 348-357.

[7] B.Gedik and L.Liu, A Customizable k-Anonymity Model for Protecting Location Privacy, In ICDCS, 2005.

[8] G.R.Hjaltason and H.Samet, Distance Browsing in Spatial Databases, ACM TODS, 1999, pp. 265-318.

[9] C.-Y.Chowand M. F.Mokbel, Enabling Private Continuous Queries for Revealed User Locations, In Proc. of SSTD, 2007, pp. 258-275.

[10] M.Kim,D.Kotz, and S.Kim, Extracting a MobilityModel from Real User Traces, In Proc. IEEE INFOCOM, 2006.

[11] D.BonehandM.Franklin, Identity-based encryptionfrom the Weil pairing, SIAM J. Computing, 2003, pp. 586-615.

[12] A.R.BeresfordandF.Stajano, Mix zones: User privacy in location-aware services, In IEEE PerSec, 2004.

[13] S.Blackman, Multiple hypothesis tracking formultiple target tracking, IEEE Aerospace and ElectronicSystems Magazine, 19(1 Part 2), 2003, pp. 5-18.

[14] S.Berchtold, C.Bohm, D.Keim, and H.-P.Kriegel,On optimizing processing of nearest neighbor queries inhigh-dimensional data space, In Proc. Conf. on DatabaseTheory, 2001, pp. 435-449.

[15] B.GedikandL.Liu, Location privacy in mobilesystems: A personalized anonymizationmodel,InProceedings of the 25th IEEE ICDCS 2005, Washington, DC, 2005, pp. 620-629.

[16] R.Cheng,Y.Zhang,E.Bertino, and S.Prabhakar,Preserving user location privacy in mobile data management infrastructures, In Int. Workshop on **Privacy EnhancingTechnologies, 2006, pp. 393-412.**

[17] P.Kalins,G.Ghinita, K.Mouratidis, and D.Papadias, Preventing location-based identity inference inanonymous spatial queries, IEEE Trans. Knowl. Data Engin, 2007, pp. 1719-1733.

[18] J.VaidyaandC.Clifton, Privacy-Preserving Top-K Queries, In Proc. of ICDE, 2005, pp. 545-546.

[19] G.Ghinita,P.Kalnis,A.Khoshgozaran,C.Shahabi, and K.-L.Tan, Private Queries in Location Based Services:Anonymsarenot Necessary, In SIGMOD'08:Proceedings of the 2008 ACM SIGMOD international conference on Management of data, ACM, New York, NY, 2008, pp. 121-132.

[20] M.GruteserandX.Liu, Protecting Privacy inContinuousLocation-Tracking Applications, IEEE Security and Privacy, 2004. pp. 28-34.

[21] T.Truta and B.Vinay, Privacy protection: p-sensitivek-Anonymity property, In PDM, 2006.

[22] H.Hu and D. L.Lee, Range Nearest-Neighbor Query, IEEE TKDE, 2006, pp. 78-91.

[23] G.HjaltasonandH.Samet, Ranking in spatial databases, In Proc. 4th Int. Symp. on Large Spatial Databases, 1995,pp. 83-95.

[24] G.Aggarwal,N.Mishra, and B.Pinkas,SecureComputation of the k th-Ranked Element, In Proc. of Int.Conference on the Theory and Applications ofCryptographic Techniques (EUROCRYPT), 2004,pp. 40-55.

[25] M. L.Yiu,C.Jensen, X.Huang, and H.Lu, Spacetwist: Managing the trade-offs among location privacy,query performance, and query accuracy in mobile services.In Proceedings of the International Conference on Data Engineering (ICDE), 2008.

[26] M. F.Mokbel,C. Y.Chow, and W. G.Aref, The New Casper: Query Processing for Location Services withoutCompromising Privacy, In Proc. of VLDB, 2006.

[27] M. Gruteser and D. Grunwald, Anonymous usage of location-basedservices through spatial and temporal cloaking. In MobiSys, 2003.

[28] ROUSSOPOULOS, N., KELLEY, S., AND VINCENT, F. 1995. Nearest neighbor queries. In Proceedings of the 1995 ACM SIGMOD Conference on Management of Data (San Jose, Calif., May 23–25). ACM, New York, pp. 71–79

[29] A. Meyerson and R. Williams, "On the Complexity of Optimal Kanonymity," in Proc. of ACM PODS, 2004, pp. 223–228.

[30] B. Moon, H. Jagadish, and C. Faloutsos, "Analysis of the Clustering Properties of the Hilbert Space-Filling Curve," IEEE TKDE, vol. 13,no. 1, pp. 124–141