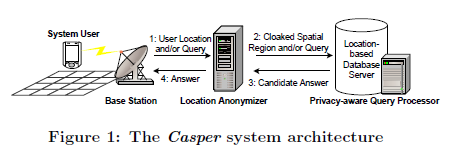
**Casper算法**

1. 开发背景

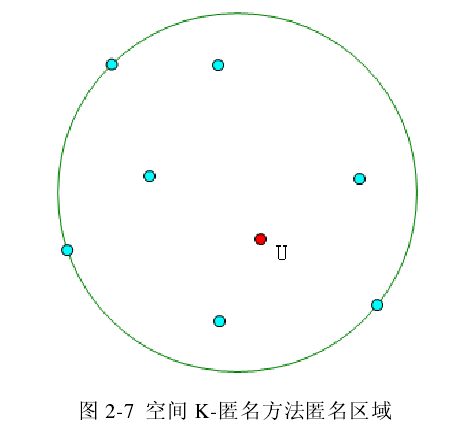
基于位置的服务（LBS， location based services）又称定位的服务，LBS通过一组定位技术获得移动终端的位置信息，提供给移动用户本人或他人以及信号系统，实现各种与位置相关的业务，如百度地图导航、美团、各种打车软件等。然而，用户在使用这些服务的同时，也面临着个人隐私信息泄漏的风险。本课题结合所学专业理论和信息安全相关知识，设计和实现一个带隐私保护的位置信息收集（提交）算法，使得用户在不泄露个人隐私的前提下提交信息并能保证数据处理的准确性。

1. 体系结构

采用第三方中心服务器体系结构。这种体系结构由移动用户、位置匿名服务器和LBS服务器三部分组成。位置匿名服务器作为移动用户和LBS服务器都信任的第三方中心服务器，是该体系结构下的核心部分（也是我们要实现的部分）。

匿名过程：移动用户在位置匿名服务器处注册等级以获得隐私保护服务，注册时提交位置隐私保护参数（一般情况下为最小匿名度K和最小匿名区域面积A）。当移动用户发送LBS查询请求时，并不是把查询请求直接发送LBS服务器，而是先发送给位置匿名服务器。位置匿名服务器收到用户的LBS查询请求后，根据用户注册时提交的位置隐私保护参数采用合适的匿名算法（本篇采用casper匿名算法）进行匿名保护，把得到的匿名区域连同LBS查询内容发送给LBS服务器。

1. 匿名技术



采用空间匿名方法。基本思想是把真实准确的用户位置信息模糊化，用包含用户真实位置的一块区域代替用户真实位置发送给LBS服务器。

空间匿名方法多采用K-匿名思想。将位置匿名服务器所覆盖的区域看成一个网络，

该区域中的一个用户即是网络中的一个节点，每个有隐私保护需求的节点需要在位置匿名服务器注册，注册内容包括最小隐私匿名度K、最小匿名区域空间A。当移动用户需要请求LBS服务时，他把用户标识符、LBS查询内容、精确的位置信息发送给位置匿名服务器，位置匿名服务器收到用户消息后，开始进行位置隐私处理。

如图2-7，它根据目标用户的精确地理位置搜索用户附近的K-1个其他用户以包含这K个用户的最小区域作为匿名区域，区域一般为圆形或矩形。如果这个用户组成的匿名区域面积不超过用户定义的最小匿名区域空间A，则此次匿名区域检索失败，如果找到满足K和A的要求，则该消息的位置隐私保护成功。接着，位置匿名服务器在消息中去除用户标识符，把查询内容连同匿名区域发送给LBS服务器。

1. Casper模型

Casper模型采用第三方中心服务器体系结构综合空间匿名方法和K-匿名思想。提出空间网格化的方法。

4.1 casper模型简介

Casper模型由位置匿名服务器location anonymizer和隐私查询处理器pricavy-aware query processor（不需做查询处理）组成。其中位置匿名服务器负责为移动用户提供位置隐私保护（只做这个）和进行LBS查询结果候选集合过滤工作（不需过滤），是独立的可信任第三方平台。

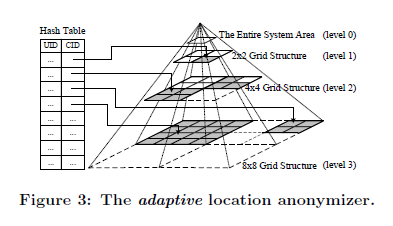
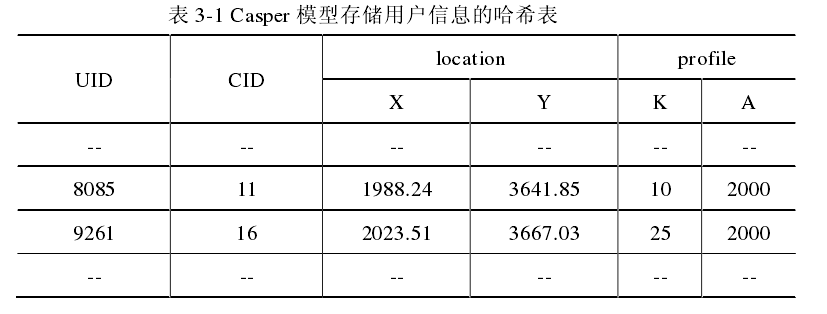


图3是Casper模型中位置匿名服务器的哈希表Hash table（不用做哈希表）和网格结构示意图。Casper模型下的位置匿名服务器将其覆盖的整个区域划分为一个个的网格cell，类似一个四叉树的结构，使用哈斯表来存储注册的移动用户的各项信息。当用户第一次在位置匿名器中注册时，位置匿名服务器需要验证该用户是否是LBS服务的合法用户。若验证通过，用户需要向位置匿名服务器发送当前位置信息和隐私保护需求，否则用户被视为非法用户，注册失败。



哈希表结构如表3-1所示，哈希表的每条记录以（UID,CID,location,profile）格式排列。其中，UID代表移动用户的身份标识；CID代表网格区域中最底层单元格编号；location表示的是移动用户的地理位置信息；profile表示移动用户的隐私需求属性，K代表一名区域的最小匿名度（即最小用户个数），A表示匿名区域的最小面积。当移动用户更新位置信息时，位置匿名服务器根据移动用户的UID找到其在哈希表中的记录，更新记录信息。当移动用户超出规定的时间与位置匿名服务器没有进行无线通信联系，位置匿名服务器则认为该用户已经离线，相对应的，在哈希表中删除该用户的记录。

4.2 Casper模型匿名算法

Casper模型的位置匿名服务器收到用户的LBS请求后，为了保护用户的位置隐私安全，需要对请求进行匿名化处理，用模糊化的匿名区域代替真实准确的用户位置信息。

匿名算法的递归搜索过程如下：

1. 先对用户节点当前所在单元格进行检查，查看该单元格内节点书目是否不小于K并且单元格面积不小于A，若二者都满足要求，则该单元格可以作为一个匿名区域，否则转到（2）；
2. 考察与当前单元格在水平向右相邻的单元格，与之合在一起作为一个匿名区域，此时区域节点数目为两个单元格节点数之和，区域面积为单个单元格的2倍，若满足匿名要求，可以作为一个匿名区域，否则转到（3）；
3. 考察与当前单元格在竖直向上相邻的单元格，与（2）区域合在一起作为一个匿名区域，此时区域节点数目为三个单元格节点数之和，区域面积为单个单元格的3倍，若满足匿名要求，可以作为一个匿名区域，否则转到（4）；
4. 若当前单元格存在父单元格（即垂直方向上的上层区域对应单元格），将当前单元格编号替换为它的父单元格编号，函数递归执行，否则返回标记-1，表示没有合适的匿名区域。

进一步解释：如上图四个正方形，从左到右从上到下依次编号1、2、3、4.假设当前节点真实位置在3单元格（即左下正方形）。

（1）若3不满足K和A，则下一步，满足则作为匿名区域；

（2）考察3和4，若3+4不满足K和A，则下一步，满足则作为匿名区域；

（3）考察3和4和1，若3+4+1不满足K和A，则下一步，满足则作为匿名区域；

（4）若当前3单元格存在父单元格，则将考察区域变为父单元格，继续从（1）递归，若不存在判断K\A优先输出合适CID。

注释：（1）上层网格和下层网格存在父子关系，由于每层网格细分粒度的不同，上层一个单元格的面积=下层4个单元格的面积之和。

1. 执行匿名算法的时候，搜索匿名区域自底向上递归，最大循环次数为（N为网格细分的粒度）。
2. Casper模型的位置隐私服务器维持一个移动用户隐私需求列表，包括最小匿名度K和最小匿名区域面积A两项，用户注册登记时可以自定义位置隐私保护需求。移动用户的隐私保护需求可以随时更改，位置匿名服务器的匿名算法根据最新更改的位置隐私保护需求提供个性化匿名保护。
3. 需求分析

5.1语言

Matlab或C++或JAVA

5.2平台

MATLAB或eclipse或DEV C