针对基于位置的服务中的隐私感知用户的统计攻击

摘要

移动设备和社交网络的融合正在刺激更广泛地使用基于位置的服务（LBS），并使其成为我们日常生活中的重要组成部分。但是，隐私泄露的问题已成为阻碍LBS进一步发展的主要因素。当LBS用户发送查询时到LBS服务器，在身份和位置方面的用户的个人隐私可能泄漏到袭击者。为了保护用户的隐私，Niu et al。提出了一种名为enhanced-Dummy Location的算法选择（en-DLS）。在本文中，我们介绍对en-DLS的两种攻击，即长期统计攻击（LSA）和区域统计攻击（RSA）。在所提出的攻击中，攻击者可以获得隐私通过分析LBS历史数据来确定用户的内容，这使得en-DLS对于用户的隐私是无效的保护。此外，本文提出了一套针对LSA和的隐私保护方案RSA。对于LSA，我们提出两个保护方法命名为多用户名（MNAME）和同一用户名称（SNAME）。为了解决地区隐私问题，我们将地图分成不同的地区隐私保护要求。为此，四级保护要求（PLs）是并且通过从虚拟变量分配一定数量的位置来保护真实位置根据位置的PL。性能分析和仿真结果表明我们提出的方法可以完全避免en-DLS对LSA和RSA的脆弱性，并且带来边际增加通信开销和计算成本。

1. 介绍

随着移动计算和网络技术的发展，手机已经成为人们生活中的必需品。除了满足dail通信的需要，手机还为人类提供了许多便利的服务[1]。 随着智能手机的发展，全球定位系统（GPS）已经在大多数智能手机上得到巩固，并为移动服务提供商提供定位智能手机的功能。 美国E-911文件指出，移动服务提供商应在125米处提供位置识别服务，以便手机的所有者在遇到火灾或劫持等危险时能够及时获得抢救和帮助[2]。

最近，许多基于位置的服务（LBS）应用成为现实。 尽管LBS带来了各种好处，由于无线网络的开放性[3]，内在的隐私泄漏问题不能被忽略。 目前，隐私泄露问题成为LBS服务广泛应用的主要障碍。

如[4]中所述，位置隐私是“防止他人获得用户的当前或过去位置的能力”。 为了在LBS服务中保护隐私，存在几个挑战：

* + 高精度：用户的身份和位置应该受到保护。 同时，应确保LBS业务的精度。
  + 低开销：用户终端的通信，计算和存储能力有限。 因此，通信开销，计算成本和存储开销在保持用户的隐私方面应该低。
  + 隐私：LBS服务器本身可能是攻击者。 它可以直接获取用户的真实位置和历史数据。

根据所使用的技术，保护用户隐私的主要解决方案可以分为混淆和匿名。 匿名，使用dummy是高效的，因为它不需要可信的第三方来保护隐私，它吸引了许多学者的“注意。 其中，Niu et al。 基于用户从历史记录中的位置发送LBS查询的概率，提出了增强型Dummy位置选择（en-DLS）[5]。 它解决了在单个LBS查询中的隐私泄漏的问题。 在[6]，Niu et al。 提出了使用缓存来提高用户隐私的缓存感知虚拟选择算法（CaDSA）和增强型CaDSA。 En-DLS具有以下特性：

* 辅助信息：在en-DLS中，辅助信息指城市中的地形信息。 虚拟变量不是从城市中的河流或山脉中选择的，而是基于位置中的历史查询概率来仔细选择的。 解决了由辅助信息导致的保护减少的问题。
* 伪装区域：为了克服k匿名性的缺点，在en-DLS中，伪装的覆盖区域被选择为尽可能大。
* 实现问题：在en-DLS中，完全考虑访问历史查询。 提出了基于接入点（AP）的方法。 通信开销是关系。

虽然en-DLS解决了LBS查询中隐私泄露的问题，但它具有漏洞。在本文中，我们介绍对en-DLS的两种攻击，即长期统计攻击（LSA）和区域统计攻击（RSA）。攻击者可以使用历史统计信息获取用户的隐私内容。对于攻击者，在损害LBS服务器之后，他可以获得大量的历史数据。我们引入一个名为LSA的攻击，以使用这些历史数据获取用户的真实身份和位置。我们研究基于LSA并提出两种方法来保护名为多用户名（MNAME）和相同用户名（SNAME）的隐私。除了LSA，攻击者可以从特定区域获取历史LBS应用。此外，攻击者

可以通过统计从该区域获取关于用户的大量信息。对于这个问题，我们提出了一种将地图中的区域划分为不同隐私级别（PL）的方法。然后，我们删除en-DLS中的一些虚拟位置，并从高PL区域中选择一些位置以保护区域的隐私。我们采取entxu2007防止bbas2013合并作为度量来分析所提出的方法的能力。性能分析和仿真结果表明，所提出的方法可以有效地保护用户对LSA和RSA的隐私。本文的主要贡献包括以下几个方面：

* 根据大多数用户的活动，我们介绍LSA。 对于攻击，我们给出了两种方法来保护用户的隐私。
* 我们根据隐私要求将地图中的区域划分为不同的PL。 我们给出一个算法，使高PL区域为高速率的虚拟区域，低PL区域低速率。 解决了攻击者通过分析历史数据可以侵犯特定区域的隐私的问题。•
* 我们分析通过熵保护用户隐私的能力。 讨论了用户LBS查询的频率，地图中的区域划分以及收集的历史信息的间隔的长度之间的关系。

本文的其余部分组织如下。 第2节给出了本文的一些初步和动机。 在本节中，我们给出LSA和RSA。 在第3节，我们提出了抵抗LSA和RSA的方法。 在第4节，我们讨论所提出的方法的安全性和性能。 第5节介绍了仿真。 在第6节中，我们回顾了相关工作。 结论和未来的工作在第7节。

1. 初步

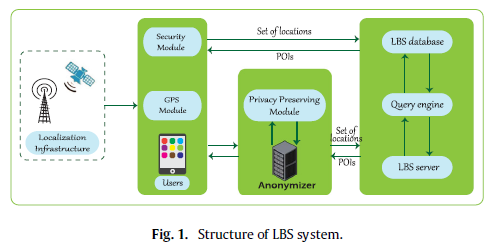
在本节中，我们首先介绍隐私度量和攻击模型。 然后，我们给出我们的解决方案的动机。

隐私权指标

为了衡量保护隐私的能力，我们需要一些指标。目前有5种度量[7]。它们是基于不确定性的度量，“基于聚类误差”的度量，基于可跟踪性的度量，k-匿名度量和基于失真的度量。在本文中，我们使用基于不确定性的度量来度量通信系统中的隐私。在[8]中，作者提出通过区分真实位置和匿名集来测量攻击者的能力。作者指出k-如果攻击者不能在同一传输中区分真实位置和k-1位置，那么真正实现匿名。在[5]中，作者提出，测量k匿名的隐私保护能力的直接方法是使用k。较大的k表示保持隐私的较高能力。然而，在该测量中存在一些缺点。例如，可以在河流，湖泊，山脉中或在中选择k-1个假人不可能的位置在路径中达到限速的速度。攻击者可以容易地将它们区分为从真实位置起的不太可能的LBS查询位置。因此，简单地使用k作为度量不能准确地表达隐私保护的能力。除了k，熵被广泛用于测量能力[5,6,8-10]。熵首先用于测量隐私[11]。众所周知，熵通常用于测量系统的不确定性。在隐私保护中，熵可以用于测量属于用户的位置的不确定性程度。在k匿名中，从攻击者的角度来看，在匿名集中包括真实位置和k-1个假人，位置是真实位置的概率是pi。在匿名集中，所有概率pi的和为1。因此，识别候选集中的真实位置的熵H是



当集合中的所有k个位置具有相同的概率时，实现最大熵，其中对于所有位置的概率pi为1 / k，并且H的最大值为log2 k。



2.2 相对模型

在本文中，我们假设用户对LBS的访问是零星的，这意味着两个连续的LBS应用之间的周期不能被忽略。 如[5]，在本文中，我们假设用户的应用信息（x，y），i，r，其他的格式。 其中，（x，y）指用户的位置。 I表示用户的兴趣，即LBS请求的类型。 范围为r。 其他包括用户的身份和其他信息。

攻击者的目标是获得用户的隐私信息，包括姓名，兴趣和位置。 攻击者可以监控周围以获取用户的应用分组，并获得用户发送的私有信息。 对手也可以监视用户以破解由LBS服务器发送给用户的兴趣点（POI）。 然后，他可以推断用户的身份，位置，兴趣等。攻击者也可以直接损害LBS服务器以获得用户的历史数据。 在本文中，我们假设LBS是攻击者。 为了商业目的，他试图获得与用户隐私相关的信息。 他对用户的实际位置和LBS查询类型感兴趣。 他不仅可以获得当前用户的LBS查询，还可以获取用户的历史数据。 他也知道隐私保护机制。

2.3 动机和新的方法

当用户通过他的智能手机发送LBS查询时，智能手机的位置首先由GPS服务确定。 然后，智能手机直接或间接地将用户的身份，位置，兴趣和查询的范围转发到LBS服务器通过匿名化。 最后，LBS服务器将根据用户的查询和反馈POI进行回复，如图1所示.

如上所述，攻击者可以获得与LBS服务器相同的用户信息。在传统的隐私保护方法中，k-1个假人被随机选择以混淆攻击者并保护真实位置。然而，在[5]中，作者发现由于地形和生活区的布局不同，在区域应用LBS服务的概率是不同的。例如，在一些城市，有河流或山脉。用户几乎不能在这些地区申请LBS服务。所以传统选择k-1个伪位置的方法由于边信息而不能有效地保护实际位置。例如，当k为20时，如果在应用LBS服务的概率低的区域中选择19个虚拟位置中的14个虚拟，则攻击者可以轻松地过滤掉14个假人。在[5]中，作者使用熵和提出的DLS来量化这个问题，其从具有相同或相似概率的网格中选择k-1个假人。 DLS算法如图2所示。

作者将城市中的地区简化为网格。 基于历史数据，作者在网格中使用不同的阴影来表示用户在区域中应用LBS服务的不同概率。 空白方块表示用户从不或很少从过去的区域发送LBS查询。 在该图中，候选单元被选择为具有与实际位置相似的概率的虚拟单元用√标记。

作者将城市中的地区简化为网格。 基于历史数据，作者在网格中使用不同的阴影来表示用户在区域中应用LBS服务的不同概率。 空白方块表示用户从不或很少从过去的区域发送LBS查询。 在该图中，候选单元被选择为具有与实际位置相似的概率的虚拟单元用√标记。

基于了en-DLS，使伪装的伪装面积尽可能大。

虽然en-DLS解决了LBS查询中隐私泄露的问题，但它具有以下漏洞。 参见图3 ，假设Bob经常在他的家里发送LBS查询。 由于历史数据的累积，真实位置将被公开给攻击者。 这是因为在en-DLS中，对于每个LBS应用程序，虚拟变量是分散的，而真实位置相对集中。 攻击者可以从历史数据中获取用户的真实位置。 此外，他可以获得关于用户的隐私内容。

在en-DLS中，假设k是20。 1为19.当用户应用LBS时，将仔细选择19个虚拟位置以保护实际位置。这种方法可以有效地保护用户的隐私在一个服务。然而，考虑到人们日常生活中的行为模式，用户停留的大多数活动场所主要集中在家庭，工作场所和固定娱乐场所（例如固定电影院或咖啡馆）。因此，提出如下的攻击。攻击者首先捕获和分析用户的历史数据。通过在特定时间段内查找特定用户的LBS请求，即使每个请求受到19个仔细选择的虚拟位置的保护，攻击者仍然能够通过分析历史数据来获得LBS请求的主要位置在一定程度上累积了隐私数据的量。然后，攻击者可以从获得的数据中推断出用户的身份，工作地点，个人兴趣和其他隐私内容。

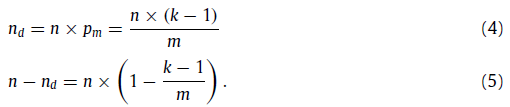
假设攻击者已经获得了用户S的历史LBS应用数据。数据中第一个LBS请求到S的最后一个LBS请求的时间间隔为t。 S以q的频率发送LBS查询。 地图分为r×r个网格。 en-DLS采用[5]。 每次用户应用LBS服务时，将存在k-1个选定的虚拟对象。 为了简单起见，我们假设用户仅在家中应用LBS服务。 假设存在具有与S的位置相同的概率并且可以被选择的m个网格作为虚拟。 然后有以下关系。 每次S发送LBS查询，m个网格中的任何一个具有概率pm为a虚拟位置。

... 

在时间间隔t中，用户S已经应用LBS服务n次。

在这些应用中，生成k×n个位置。 在位置中，真实位置的数量是n，位置是用户的家。 其余的（k-1）×n个位置是虚置。 在m中，每个网格在（k-1）×n中具有nd个时间为虚

拟。



我们假设每当用户S应用LBS服务时，有足够的网格是虚拟的。 也就是说，误差远大于k-1。 假设k-1> 0。我们可以得出pm <1和nd <n的结论。 计算后，得到：



我们得出以下结论：m越大，n-nd越多。也就是说，只要m足够大，m个虚拟位置的每个网格的时间将接近0.因为n是用户S应用LBS服务的次数，并且它是常数。因此，只要n足够大（也就是说，用户的LBS应用的历史数据就足够了），我们可以得出结论，LBS应用在实际位置的时间将远远大于任何虚拟位置。这表明如果用户的真实位置相对集中，虚拟对象的位置是相对的分散因为它们的随机性和虚拟可以被攻击者忽略。因此，真正的位置不能有效地由这些假人保护。需要说明的是，在本文中，我们假定用户应用LBS服务的频率q作为确定的值。实际上，用户随机应用LBS服务方式。为了简单起见，在本文的后半部分，q也将被认为是确定的。对随机q的现实分析可以是未来的工作。在上述方法中，攻击者可以通过分析历史数据获得隐私内容。在本文中，我们称之为这种攻击长期统计攻击（LSA）。

下面，我们将通过实验说明攻击。

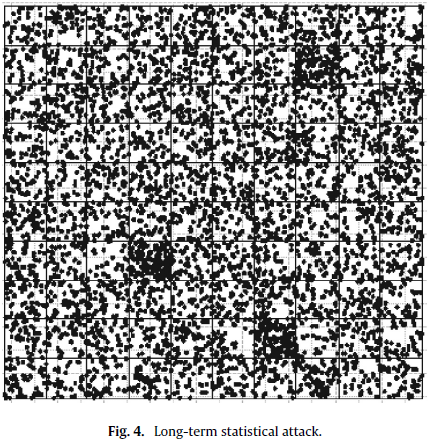
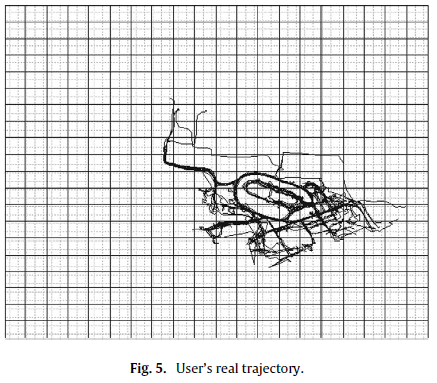
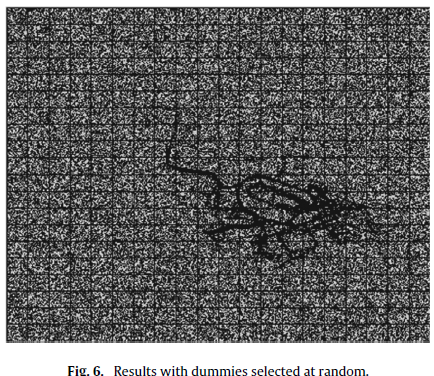


图4是在一个月中应用LBS服务的用户Bob的统计图。 Bob每天应用LBS服务约3次。 地图分为10×10大小的网格。 网格中的每个点表示LBS服务器从中接收到一个LBS应用程序位置。 在这些应用程序中，一些应用程序来自真实位置，其他应用程序来自选定的虚拟对象。 其中，Bob的真实LBS应用的数量是90.每当Bob发送真实应用时，将随机生成19个虚拟位置。 为了说明问题，我们在实验中随机选择虚拟变量。 根据en-DLS在[5]中选择虚拟变量的方法与此类似。

从图中我们可以看出，大多数用户的应用位置都在三个网格中。 三个网格分别对应于Bob的家，公司和咖啡馆，其中Bob花费他的闲暇时间。 根据图1。 4，攻击者可以获取Bob的真实的位置，然后推断鲍勃的隐私，如身份，工作地点和个人习惯。





图。 图5是用户在一天中的真实应用位置。 用户每30秒应用LBS服务。 图。 6是为每个真实应用程序生成19个虚拟变量的结果。 从图中，我们可以看到虽然添加了虚拟，但是用户的真正LBS应用集中的位置不能被有效地保护。

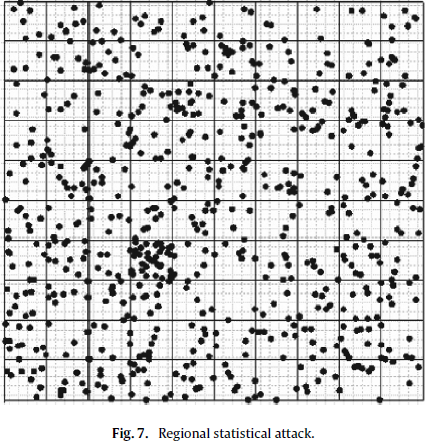
除了LSA之外，在en-DLS中还存在隐私保护中的另一个漏洞。攻击者可以根据历史数据收集和分析特定区域中的LBS查询。如果该区域对应于Bob的家并且在很长一段时间内，应用程序从这个区域主要是关于健康[12]，攻击者可以推断，该地区的用户与他的健康有问题，即使使用[5]中的方法。如果Bob独自存在，攻击者可以获得居住在该地区的用户是社交的Bob工程，他可以知道鲍勃有健康问题。攻击者可以把这个信息卖给商业机构，而鲍勃会很遗憾地发送大量的医疗保健广告。鲍勃会觉得他的隐私被侵犯。如果鲍勃与他的家人一起生活，即使它不能准确地确定谁有健康问题和无法确定用户的身份，攻击者也可以得出的结论是，有人在该家庭有健康问题。因此，家庭的隐私被侵犯。

另一个例子是在一个秘密部门（如政府机构，军工企业）有大量的员工。攻击者分析来自该区域的历史数据中的LBS查询。假设在一段时间内，下午5点，部门的大部分员工下班，他们通过LBS服务应用交通信息。即使假人被采取，通过历史数据分析，攻击者可以知道在下午5:00有许多用户申请流量服务。然后，攻击者可以得出结论，该部门在下午5:00停止工作。如果在一段时间内，攻击者从大多数用户在早上6:00发送流量查询的历史数据获得，他可以很容易地得出结论，在该时间段，该部门在夜间加班。他也可以推测，在这段时间内，该部门正在从事一个重要的秘密项目。根据类似的痕迹，攻击者可以获得与部门工作相关的更多或更少的信息。然后，部门的秘密将被违反。

另一方面，如果地图被分成许多网格，使用伪实现隐私保护可能导致敏感区域具有较小概率成为其他LBS查询的虚拟项。 因此，从该地区发送的申请将减少。 通过历史数据分析，攻击者可以很容易地从该地区获得真正的应用。

基于[5]中的方法，LBS服务器是攻击者。地图分为r×r个网格。假设在网格A中，只有用户S应用LBS服务。在时间间隔t，从攻击者的角度看，网格A中的用户S应用LBS服务s次。 s = sreal + s dummy，其中sreal表示在t的时间间隔中来自A的用户的实际LBS应用的次数，并且s dummy表示在t中作为其他LBS应用的A的时间。假设存在与A具有相同或相似概率的m个网格。也就是说，当A中的S应用LBS服务时，m个位置可用于选择为虚拟。进而，当在m个位置中的任一个中的用户应用LBS服务时，A可以是虚拟的。假设在所有这些位置中有n个用户，并且来自每个用户的LBS应用的频率是q。当用户应用LBS服务时，生成k-1个假人。然后，在t的时段中，m个位置中的总LBS应用的数量是n×t×q。在一个应用中，A为虚拟概率pA为：





在t期间，A为其他人的虚拟的总时间，sdummy，是：



我们可以从（8）看出，更大的是m，更小的是sdummy。 如果sreal足够大，来自A的LBS应用程序可以被认为是真正的应用程序。 因为LBS服务器是攻击者，他可以很容易地从（8）获得k和计算sdummy。 如果他得到在一个特定的地区，s远远大于sdummy，他可以得出结论，在该地区有大量的实际应用。 如果大多数应用程序的利益碰巧是相同的，他可以将该利益与该地区相关联。 在本文中，我们称之为这种攻击区域统计攻击（RSA）。

下面，我们将通过实验说明攻击。

图7描绘了在特定地区下午5:30的LBS应用。图中的每个点表示该位置中的真实LBS应用或虚拟。 收集更多积分的网格是一家公司。 公司的员工在5:30停止工作，他们在当时应用LBS服务的公共汽车信息。 每次发送真实应用程序时，将生成19个虚拟对象。 从图中可以看出，LBS应用主要集中在该地区。 攻击者可以分析历史数据。 如果在历史数据中，在下午5:30，关于总线信息的LBS应用程序经常从该地区发送，并且攻击者通过社会工程获得在该地区有一个公司，他可以容易地得出结论，该公司停止工作在5： 30点.

1. **增强K匿名方法**

在本节中，我们提出了防御LSA和RSA的方法。

3.1 防御LSA的方法

由于用户的历史数据可以被攻击者用于LSA，我们提出两种方法来抵抗攻击。

3.1.1名称

在此方法中，用户存储多个用户名。 每次应用LBS服务时，用户从设置为当前用户名的用户名中选择一个名称，并将其发送到LBS服务器。 在生成这些名称的过程中，应避免使用不常见的用户名。 这种方法的优点是隐私保护不依赖于第三方。 由于来自用户的LBS应用是零星的，攻击者很难将两个用户名链接到一个用户。 因此，不需要混合区。 这种方法可以有效地保留隐私与虚拟选择方法相结合在[5]。

3.1.2 SNAME

在这种方法中，每次用户应用LBS服务时，查询由匿名者改变并发送到LBS服务器匿名者。 匿名程序将每个用户名更改为相同一个，然后将查询发送到LBS服务器。 假设的个数匿名者的用户是m。 只要m足够大，LSA就可以被捍卫。

3.2 防御RSA的方法

我们提出了一种防御RSA的方法。

我们根据网格的隐私要求将地图中的区域划分为4个PL。

一级区域主要包括一些秘密部门。 在这种地区，隐私的泄露可能会侵犯部门的利益。

第二级区域主要包括可以公开用户个人隐私的区域，例如某人的住所等。从这种区域的LBS应用中，攻击者可以容易地确认用户的身份和与其相关的信息该区域。

第三级区域主要包括人少的区域。 在这种地区，隐私容易受到侵犯。

第四级区域主要包括有更多人的区域，例如商业区等。在这种类型的区域中，隐私泄漏的可能性较小。

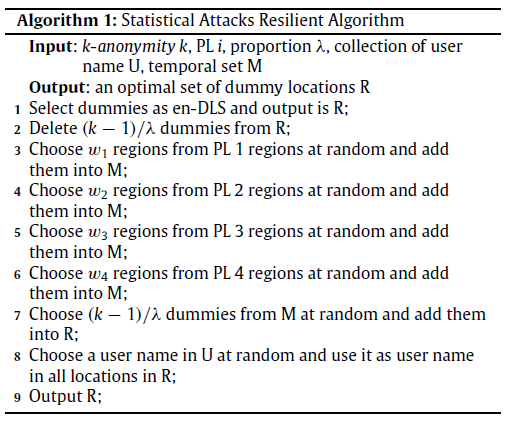
对于第三级区域，我们将通过一个例子说明问题。 在一个遥远的工厂，有两个员工。 一个员工是攻击者，他可以获得该地区的LBS历史数据。 虽然[5]中的en-DLS具有较小的概率选择该区域作为其他人的虚拟，如果LBS应用经常从该区域发送并且应用的兴趣相同，则攻击者可以断定这些LBS应用来自其他雇员，并且他可以获得应用的兴趣类型 。 因此，在人少的地区，隐私泄露的可能性大于人多的地区。

相反，对于第四级，因为在该区域中存在更多的LBS用户，所以实际身份和位置将由同时应用LBS服务的LBS用户很好地保护。 因此，在经常有大量的区域的LBS用户，隐私泄露的概率较小。

对于上述四级区域，基于en-DLS，我们提出了抵抗RSA的方法。 在k-1个虚拟变量中，我们根据一定比例删除一些虚拟变量。 我们使第一级区域具有高概率作为虚拟，而第四级区域具有低概率。 优点是，将有大量的LBS查询来自高隐私要求的区域。 然后，攻击者无法区分哪一个是真正的应用程序。 其余的虚拟对象由en-DLS选择。

算法如下。

在算法1中，首先根据en-DLS选择k-1个虚拟位置。 其次，在这些k-1个假人中，我们删除了一些假人作为1 /λ的比例。 然后我们构造区域匿名集M.首先，我们从PL 1随机选择w1个区域区域并将其放入M.第二，我们从PL 2区域随机选择w 2个区域并将它们放在M. 第三，从PL 3区域随机选择w 3个区域并将它们放入M.第四，从PL 4区域随机选择w 4个区域并将它们放入M.典型的比例是两个= 4和w 1：w 2：w 3：w 4 = 8：6：4：1。 然后，我们从M中选择（k。1）/两个虚拟变量，并将它们添加到集合R中。接下来，从用户名U中随机选择一个名称，并将用户名与虚拟变量组合。 最后，k个位置将被发送到LBS服务器。 在该算法中，采用MNAME来抵抗LSA。



1. 分析结果
   1. 安全分析

在我们的方法中，对用户和LBS服务器之间的通信（例如窃听）的攻击可以通过加密来防御。 接下来，我们主要关注LBS服务器是攻击者的攻击。 他希望获得一些内容通过使用历史数据隐私。

熵是衡量隐私保护能力的有效指标。 在本节中，我们使用熵来测量所提出的方法的隐私保护能力。

首先，我们分析在抵抗LSA中使用MNAME的方法。 假设系统中的每个用户存储m个用户名。 用户S还存储m个用户名。 在每个LBS查询中，用户使用用户名的概率是ps。



假设S使用用户名u发送LBS查询，并且在相同LBS服务中有r个用户在除了S之外的用户名集合中具有U.假设来自这些r个用户的LBS应用的频率是q。 假设来自S的LBS应用和来自S的下一个应用之间的时间间隔为t。 然后在t中，具有用户名u的每个用户已经应用了q×t次的LBS服务。 在这些应用中，使用u作为用户名的应用的次数是ps×q×t。 对于每个应用程序，将生成k - 1个假人。 然后在t中，u用作应用程序的用户名对于k×r×ps×q×t + k次（包括来自S的应用）。 其中，S的实际应用是1倍。 也就是说，在两个应用程序之间，攻击者将S的真实位置与匿名时间区分开的熵t的周期为：



如果用户不使用不常见的用户名，r将很大。 用户应用的频率越高，q越大，实际位置将被有效地保护以防止LSA。

下面，我们将讨论使用SNAME防御LSA的问题。 假设S是应用LBS服务的用户。 假设有r个用户使用相同的匿名者，这些r个用户的应用频率为q。 假设来自S的LBS应用和来自S的下一个应用之间的时间间隔为t。 然后，在t的时间间隔中，使用相同匿名者的每个用户应用LBS服务q×t次。 这些用户的用户名与S相同，为u，并且在每个应用中，将生成k-1个虚拟位置。 然后，在t中，使用u作为用户名的应用的数量是k×r×q×t + k（包括S的应用）。 其中，S的实际应用只有1次。 也就是说，在两个应用程序之间，攻击者在t的时间段中将S的真实位置与匿名集合区分开的熵是：



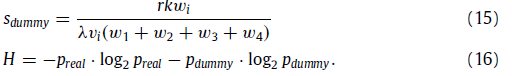
最后，我们将讨论根据算法1防御RSA的方法的安全性。假设在网格A中，只有S应用LBS服务，并且来自S的应用与来自S的下一个应用之间的时间间隔是t。 假设在t，r个用户应用LBS服务。 对于每个用户的应用，k。 生成1个虚拟变量。 在k。 穆米斯角 k /两个虚拟变量根据en-DLS生成。 剩余的k /两个虚拟变量用于保持高PL区域的隐私。 Letw = w1 + w2 + w3 + w4并假定A的PL是i。 在图中的PL i区域的数量是vi。 根据算法1，wi区域将从vi中随机选择并添加到M.因此，将A添加到M中的概率pSM是：



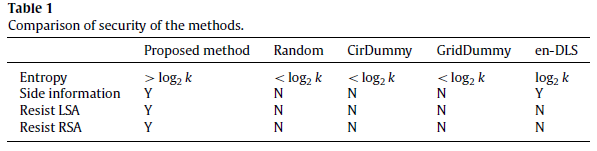
令g = k /λ。 在M中，A被选择为a的概率虚拟位置，pSD为：

根据这些，我们可以在一个应用中得出结论：A被选择为其他应用的虚拟位置的概率pSR是：

在S的两个应用程序之间，有r个用户应用LBS服务。 然后，区域A成为其他应用程序的虚拟对象的时间，sdummy是：



根据第2节中的分析，我们可以知道对于LBS服务器，在来自区域A的应用程序中，s = sreal + sdummy。 该区域的隐私取决于虚拟。 因此，wecan通过确保sdummy比sreal更保守隐私。 在公式（15）中，sdummy与vi成反比并与r成比例。 更大的是wi，更大的是sdummy。 在实践中，我们可以通过调整wi和vi来保护区域的隐私。 vi应该小，wi应该大。 也就是说，具有高PL的区域将被给予更多的保护，例如一些秘密部门和人民的重要居住。 另一方面，随着LBS业务的普及，LBS应用的频率将越来越高。 因此，可以确保大的r。 因此，本文提出的方法可以解决随着LBS的普及，RSA的问题。



在表1中，我们将所提出的方法的熵与随机选择伪随机的方法以及[5,13]中的方法进行比较。

在表1中，只有本文中的方法考虑LSA和RSA。 因此，从统计学的角度来看，本文提出的方法的熵大于log2 k。 随机选择dummy的方法的熵小于log2 k，因为该方法在选择虚拟变量时具有脆弱性。 对于CirDummy和GridDummy，当考虑辅助信息时，熵低于log2 k。 对于en-DLS，当不考虑统计攻击时，熵为log 2 k，而当考虑统计攻击时，熵小于log 2 k。

* 1. 性能分析

在下文中，我们将分析所提出的方法的性能，并表明它们具有很强的实用性。 我们比较所提出的方法与其他方法的性能。

* + 1. 效用

在我们的方法中，对于每个LBS查询，真实位置与虚拟对象一起发送。 在MNAME和SNAME中，只有用户的身份被更改。 位置的准确性没有受到影响。 因此，所报告的位置仍然可以向用户提供合理的查询答案。 在算法1中，仅改变虚拟位置。 实际位置的精度没有受到影响。 所以，LBS服务器仍然会给出合理的答案。 总之，所提出的方法的效用仍然是合理的en-DLS。

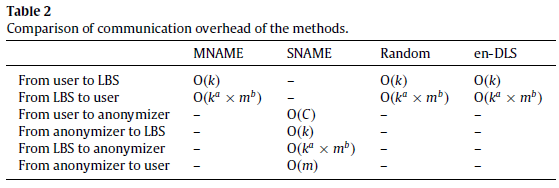
* + 1. 通信开销

与其他方案相比，本文提出的MNAME不会带来额外的通信开销。 从用户发送的LBS应用包括实际位置和k-1个假人，并且通信开销是O（k）。 从LBS服务器发送的信息的通信开销与返回的POI和k。 假设返回POI为m。 那么LBS服务器的通信开销为O（ka×mb），其中a和b是常数.

在SNAME中，需要一个匿名程序。 通信开销分为从用户到匿名器的开销，从匿名器到LBS服务器的开销，从LBS服务器到匿名器的开销以及从匿名器到用户的开销。 在其中，从用户到匿名器的通信开销仅包括加密的真实用户名，位置，兴趣等。因此，开销是常数。 从匿名器到LBS服务器的通信开销包括真实位置和用户名改变的k-1个虚拟变量。 开销是O（k）。 从LBS服务器到匿名器的通信开销与返回的POI和k相关。 它是O（ka×mb），其中a和b是常数。 的从匿名器到用户的通信开销仅包括用于真实应用的POI。 所以它是O（m）。

在防御RSA的方法中，一些伪装占据了原始伪装的位置以为高PL区域提供隐私保护。 因此，不需要额外的通信。 主要通信开销来自在地图中分布区域的PL。 PL的信息不经常更新。 它可以每月更新一次或每年更新一次。 因此，这部分通信开销在理论分析中可以忽略。

在表2中，我们比较了通信开销所提出的方法选择随机虚拟和en-DLS的方法。 在表中，C表示常数。



* + 1. 计算成本

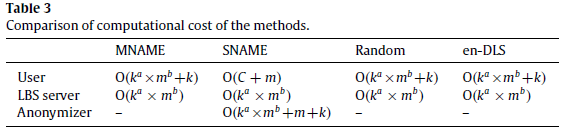
计算成本主要包括位置的加密和解密。

在MNAME中，用户随机选择用户名并生成虚拟对象。 为了防止攻击者窃听，传输的信息应该被加密，并且结果将被发送到LBS服务器。 用户的计算成本是O（k）。 在信息到达LBS服务器后，服务器对返回的POI进行加密并发送给用户。 LBS服务器的计算成本是O（ka×mb），其中a和b是常数。 用户获得加密的POI。 然后他解密POI。 计算成本为O（ka×mb）。

在SNAME中，当发送LBS查询时，用户的计算成本包括真实用户身份和真实位置的加密。 成本是O（C）。 匿名器生成k.1个虚拟位置，并将结果发送到LBS服务器。 计算成本是O（k）。 LBS服务器获取应用后，生成POI。 然后，将POI加密并发送到匿名器。 LBS服务器的计算成本为O（ka亊mb）。 匿名器解密POI。 然后匿名程序加密结果并将其发送给用户。 计算成本为O（ka亊mb + m）。 在用户接收到m个加密的POI之后，他对它们进行解密并获得最后的结果。 计算成本为O（m）。 因此，用户的总计算成本是O（C + m）。 匿名者的总成本是O（ka亊mb + m + k）。

在防御RSA的方法中，原始假人仅被根据PL选择的假人替换。 因此，仅添加选择虚拟变量的计算。 由于选择假人的计算成本非常小，这在性能分析中忽略计算成本。

在表3中，比较了方法的计算成本。



* + 1. 存储开销

在MNAME中，对于移动设备，将添加用户名集的存储。 存储开销与存储的用户名的编号有关。 它是O（n），其中n是存储的用户名的编号。 在SNAME中，由于使用匿名程序，额外的存储移动用户不需要开销。 匿名者的存储开销与应用LBS服务的用户数量有关。 在防御RSA的方法中，用户需要在地图中获取和存储PL。 存储开销为O（r×r）。 由于PL不会经常被更新，所以存储开销是常数。

1. 实验和结果

下面，我们将通过实验验证所提出的方法的性能。 实验重点在于方法的隐私保护能力，用户的通信开销和用户的计算成本。

5.1 实验设置

我们做实验的机器是一台带有Intel Pentium CPU G630 2.7 GHz和8.0 GB RAM的计算机。 电脑的操作系统是Win7 64bit。 我们的实验是在Matlab 2012b。 在我们的实验中，地图被划分为网格1000亊1000。 10 000个用户随机分布到网格中。 地图中的POI总共是1000。

5.2 模拟结果

5.2.1 熵

在MNAME中，所有用户有1000个用户名。 假设每个用户从用户名中选择5个名称作为用户名设置。 用户LBS应用的频率为每小时1次。 。 8，我们比较我们的方法的熵与其他使用虚拟在保护隐私。

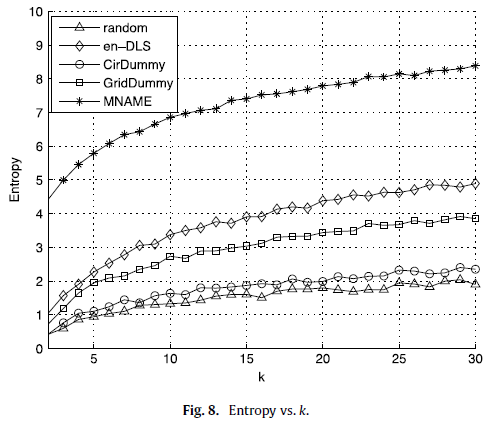
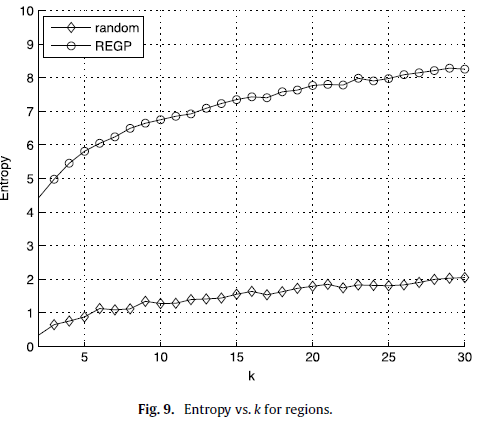


图 8，en-DLS是[5]中提出的方法。 随机是随机选择虚拟变量的方法。 CirDummy和GridDummy是在[13]中提出的虚拟圆和虚拟网格。 MNAME是本文提出的方法。 从图。 8，我们可以看到MNAME的熵是最大的。 它大于log2 k。 原因是MNAME认为LSA并使用多个用户名。 LBS查询来自使用相同用户名的其他人保护用户的真实查询。

在图中，仅调查了NAME。 SNAKE的熵与相同匿名者中的用户相关。 省略。

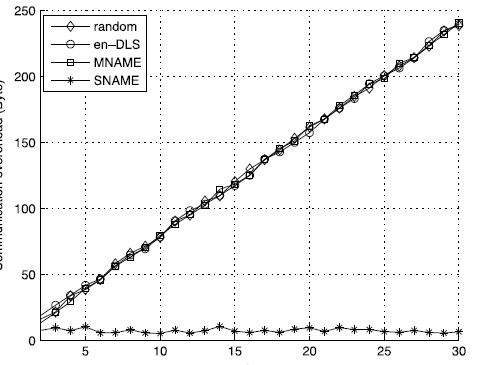
接下来，我们将研究攻击者将实际应用与特定区域区分开的熵。 假定区域的PL为1.λ= 4，w1：w2：w3：w4 = 8：6：4：1。在地图中，总共有100个PL为1的区域。

图 9，random是随机选择虚拟变量的方法。 REGP是本文提出的防御RSA的方法。 从图中可以看出，本文提出的方法具有比随机选择虚拟变量的方法更大的熵。 随机方法的熵约为REGP的24.9％。



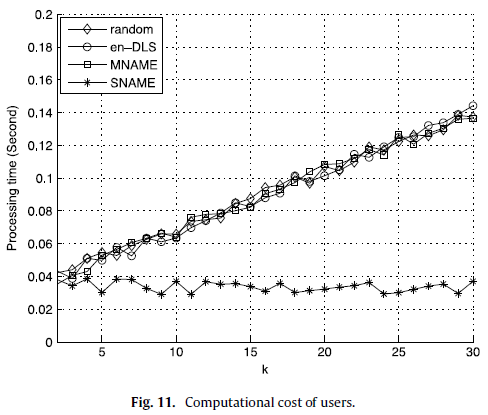
5.2.2。 通信开销图。

图10描述了k和用户的通信开销之间的关系。 发送2D位置的开销大约为8字节。 在图中，random是随机选择虚拟变量的方法。 MNAME和SNAME是本文提出的方法。 在图中，随机，en-DLS和MNAME的通信开销是相似的。 在随机，en-DLS和MNAME中，虚拟变量在用户端生成。 因此，通信开销随着k的增加而增加。 在SNAME中，用户只需要向匿名者发送真实位置。 因此，通信开销是常数。



5.2.3。 计算成本。

如图11所示，用户的计算成本主要包括位置和返回的POI的加密和解密。 在图中，随机，en-DLS和MNAME的计算成本是相似的。 由于在这三种方法中，在用户侧生成虚拟，在用户侧对信息进行加密和解密，所以计算成本高于SNAME。 随机，en-DLS和MNAME的计算成本随着k的增加而增加。 在SNAME中，采用匿名者来保护隐私。 因此，计算成本仅包括实际位置和返回的POI的加密和解密的成本。



6.相关工作

在本节中，我们回顾一些关于用户的LBS隐私保护的研究。确保用户的隐私和安全具有重要意义当向用户提供高精度定位服务时。许多解决方案已经在文献中记录。这些解决方案可根据技术分为两类用过的。他们是混淆和匿名[14]。应用LBS服务，模糊不会向LBS服务器发送真实位置，但隐藏真实的位置[15-18]。这种方法主要包括添加噪声并向LBS服务器发送类似的位置。在[14]中，提交人提出收集地点和利益从遭遇LBS用户和隐藏通过使用不同用户的历史数据来确定真实位置。虽然这种方法可以在一定程度上保护隐私由于发送位置，LBS服务的精度不高不准确。不同于混淆，匿名方法[4,19,20]旨在使用户的真实位置不能与LBS服务器发送的位置区分开，以便保护真实位置。匿名使用的主要技术方法包括伪名称，k-匿名性等。在[4]作者提出混合区。在混合区域，用户受到保护伪名称。但这种方法缺乏实用性很难在现实中应用。另一方面，k匿名是主要方法在隐私保护领域。它是首先用于保护数据库中的数据[19]。用于保护位置隐私的Gruteser和Grunwald [8]。为了实现k匿名，有两个主要方法：伪装和使用虚拟。在掩饰中，有一个匿名者。的匿名者扩展真正的位置到一个大的掩蔽地区确保在该区域至少有k个用户发送LBS查询与此同时。然后匿名者发送整个掩饰区域到LBS服务器。然而，在[21]中，作者指出伪装的弱点。另一种实现k匿名的方法是发送真正的位置与k - 1个假人，使LBS服务器不能区分哪一个是真实位置。理想情况下，每个位置的概率为真实位置是1 / k。在[22]中，作者研究了三者之间的三角权衡隐私，服务质量和带宽。笔者指出基于虚拟的位置隐私保护机制提供为给定的质量和带宽组合提供最佳保护约束。在[23]，Kido et al。使用虚拟实现k匿名。通过随机移动生成虚拟位置。的作者在[13]设计了两个选择虚拟的计划。他们是CirDummy和GridDummy。在CirDummy，虚拟位置基于虚拟圆生成，而在GridDummy中，基于虚拟网格生成虚拟位置。虚拟这些方法中的位置被仔细选择以确保遮蔽面积足够大。在[5]中，作者注意到效果的关于用户隐私披露的侧面信息，特别是在城市的地形约束。作者提出Dummy-位置选择（DLS）方案。当用户发送LBS查询时，在具有类似查询的位置中选择虚拟变量概率作为真实位置。为了增加覆盖面积DLS，en-DLS。在en-DLS中，掩蔽区域足够大同时实现大熵。

根据匿名者使用或不使用，目前的隐私保护方法可以分为可信第三方（TTP）方案和可信第三方免费（TTP）方案[24]。在TTP方案中，用户的隐私保护由可信的第三方称为匿名者。匿名者从用户处接收应用程序，并通过使用伪装或其他方法来保护应用程序。然后受保护的应用程序将由匿名器发送到LBS服务器。缺点是匿名者可能是系统的瓶颈，它可能是攻击的主要目标。一旦匿名者被盗用，使用匿名者的所有移动用户可能面临安全威胁，即单点故障。隐私使用匿名器的保存方案包括[24.28]中的方法。不使用匿名程序的方案称为无TTP方案。在这些方案中，移动设备的隐私保护不依赖于可信第三方。用户生成隐私保护信息依赖于智能移动设备自己，结果将直接发送到LBS服务器。这种方法主要包括[5,9,10,23,29,30]中的方法。

另外，根据用户向LBS服务器请求的频率，隐私保护的问题可以分为偶发的隐私保护和持续的隐私保护。 在零星的隐私保护，两个LBS查询之间的期间从用户不能被忽视。 例如，用户发送LBS查询以获得附近的餐馆地址。 下一个LBS查询可以在当前查询之后应用长时间段。 在连续的隐私保护中，用户连续发送LBS查询。 对于例如，在路上移动的汽车向LBS服务器发送连续的LBS查询以获得当前的交通信息。 在这种LBS服务中，用户的轨迹还可以揭示用户的隐私。 因此，在这种隐私保护中，不仅仅是用户位置应该被保留，而且轨迹需要被保护。 连续隐私保护的方案主要包括[18,31.34]中的方案。

7.结论

在本文中，我们首先介绍了en-DLS中存在的漏洞。攻击者可以通过使用历史数据的LSA和RSA违反用户的隐私。然后，我们给出了防御这两种攻击的方法。对于LSA，我们提出MNAME和SNAME。对于RSA，我们建议将地图中的区域划分为不同的PL。我们根据PLs替换了一些虚拟变量和区域的位置。此外，我们使用熵来分析所提出的方法的隐私保护能力。在分析中，我们给出了熵，频率的关系LBS应用程序，以及攻击者获取的历史数据的持续时间。模拟结果表明，本文提出的方法保护用户的隐私免受历史统计2007年locationardagna 2007年位置duckham 2005正式攻击具有高性能。未来的工作主要集中在用户的应用的随机性和LBS应用在一天中的不同时间的急剧变化。