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

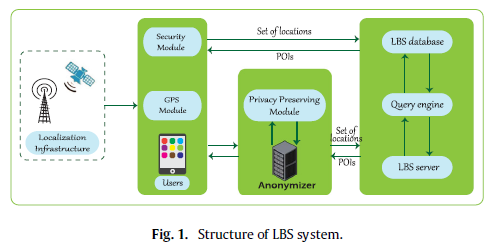
在本节中，我们首先介绍隐私度量和攻击模型。 然后，我们给出我们的解决方案的动机。

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。

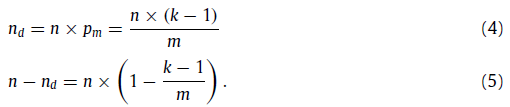
假设攻击者已经获得了用户S的历史LBS应用数据。数据中第一个LBS请求到S的最后一个LBS请求的时间间隔为t。 S以q的频率发送LBS查询。 地图分为r×r个网格。 en-DLS采用[5]。 每次用户应用LBS服务时，将存在k-1个选定的虚拟对象。 为了简单起见，我们假设用户仅在家中应用LBS服务。 假设存在具有与S的位置相同的概率并且可以被选择的m个网格作为虚拟。 然后有以下关系。 每次S发送LBS查询，m个网格中的任何一个具有概率pm为a虚拟位置。

... 

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）。