**移动商务用户的隐私保护算法**

第一作者:王玉

南京邮电大学,江苏,南京

摘要

根据移动用户对隐私信息的个性化定义，提出（K，L，P）匿名模型。基于该匿名模型，提出了用于匿名组交换和合并过程的新的隐私保护算法（AMAGS）。AMAGS可以用来保护用户的位置，标识出公路网络上移动用户的位置和其他敏感信息。最后，基于移动用户在真实的公路网上的隐私分布，可以通过实验来验证AMAGS的可行性和优势。

关键词：匿名模型；个性化隐私信息；隐私保护算法；服务框架

Abstract

According to the defined personalized privacy profile of the mobile user, a (K, L, P)-anonymity model is formally described. Based on the model, a new privacy-preserving algorithm for exchanging and merging processes for generating anonymity sets (AMAGS) is proposed. The processes of exchanging and merging are formally described. Finally, based on a real road network and generated privacy profiles of mobile users, the feasibility and advantages of AMAGS are experimentally validated.

Keywords: Anonymity model; Personalized privacy profile; Privacy-preserving algorithm; Service framework

1. 介绍

近年来，随着新的应用信息和通信技术的普及，移动电子商务得到迅速的发展。其中使用最广泛的是基于位置的移动商务应用­—移动广告。使用LBS时，移动用户通常需要向服务提供商发送其查询请求和精确的位置。服务提供商可能会收集，处理和存储用户的地理位置信息，所以与地理位置相关的隐私问题也引起了越来越多的关注。本研究试图解决以下问题：(1)为移动电子商务提供个性化的隐私保护服务框架。(2)如何在移动电子商务的背景下定义个性化的移动用户的隐私要求。(3)根据定义的匿名模式，建立一个新的隐私保护算法。

在第2节，在了解与移动电子商务中的隐私信息相关的概念和工作后，对在移动环境中常用的隐私保护技术进行了讨论。在第3节，建立了为移动电子商务提供位置服务的隐私保护服务框架。在第4节，根据移动用户定义的个性化的隐私信息建立了（K,L,P）-匿名模型。基于匿名模型，提出了一种新的隐私保护算法。在第5节，基于一个真实的路网和移动用户的隐私信息，通过实验验证AMAGS的可行性和优势。结论最后介绍。

1. 基于公路网上的隐私保护技术

大多数现有的研究结果集中在基于所述位置的K-匿名模型来保护用户的位置和身份识别码信息。K-匿名模型即将用户的精确位置用一个至少包含其它k- 1个人的空间区域来代替,从而使攻击者无法将某一个人的位置信息通过推理攻击的方式与其身份相匹配。

已经有一些关于移动电子商务的敏感信息保护的研究。TaintDroid通过第三方应用程序的流量来跟踪隐私敏感数据并扩展到Android手机平台。Pan是第一个研究公路网络中的用户敏感信息的保护技术。他们提出了（K，L，P）-匿名模型和P3RN匿名算法。这种方法支持个性化的隐私要求，确保用户的位置信息安全。根据改进该方法的主要方法，我们将提出一个新的隐私保护算法。

1. 移动电子商务的隐私保护服务框架
   1. MCA的隐私保护服务框架

在本文中，我们认为MCA是一个值得信赖的信息服务平台。MCA的隐私保护服务框架如图3所示。它有四类实体组成：MCA平台，信息服务供应商，拥有实体店的供应商，以及移动用户。MCA平台作为一个可信的提供匿名服务和全面基于位置信息服务的第三方，包括根据移动用户的查询请求和目前的位置向供应商提供推荐。



* 1. 基于MCA供应商推荐服务模式的隐私保护

服务模型如下:

* 移动用户在MCA平台上注册他们的基本信息，这些信息可以被访问。
* 当接收到查询请求时，MAC平台将执行一个基于用户位置查询的匿名算法， 产生一个满足指定匿名模式的匿名集。
* 向MCA的信息服务提供商发送匿名查询请求和匿名集。
* 各信息服务提供商使用特定的算法来处理查询，并将查询结果返回到MCA平台。
* 根据查询用户的精确位置和多个信息服务提供商的推荐供应商列表，在MCA平台上使用排列算法并发送最佳推荐列表给查询用户。
* 当用户采用该建议，它可以被看作是一种有效的建议。

根据上述的供应商推荐模型，MCA平台接收用于查询用户精确的位置信息，并生成匿名查询，而各信息服务提供商必须基于匿名集执行它自己的推荐算法并向服务商提供推荐列表。我们会根据匿名集不断的交换和合并过程为平台建立一个匿名模型并给出一个隐私保护算法。

1. 公路网络的个性化隐私保护算法

4.1 基本定义

信息敏感性是指对可能会泄露的信息的访问的控制。不同的移动用户可以基于不同的隐私倾向分配给信息不同的灵敏度水平。提交给匿名服务器的所有查询请求可以被划分为不同的灵敏度类别，以及每个类别都标有一个[0，1]之间的值来表示相应的信息灵敏度。标记值越大，敏感度级别越高。

**定义1（个性化隐私信息）**用户可以将个人隐私信息描述为四元组（k,qsr,sd,p）。

在此定义中，匿名需求（k）代表移动用户可以接受的最小匿名级别，这意味着匿名集中至少有K个移动用户；查询灵敏度需求（qsr）表示移动用户允许最大的查询灵敏度。如果查询Q的灵敏度值大于qsr，则Q对于该移动用户来说是敏感查询；段多样性要求（sd）代表匿名地区不同路段的最小数量。sd的值越大，移动用户的位置匿名的水平越高；设定灵敏度需求（p）的意味着匿名集中移动用户对敏感查询的最大容限比。p的值越大，设定的灵敏度越严格。

**定义2（匿名集）**这是由在匿名区域的移动用户组成。将匿名集设为AS，代表在匿名集AS中的移动用户的数目。

对于所有的，u的匿名需求是u.k，其查询灵敏度需求是u.qsr，其集灵敏度需求是u.p。u发送的查询灵敏度是u.qs。另外，在AS中的用户的最大匿名需求被表示为max u.k，匿名集中道路段的数量被表示为AS.Count\_S，u的敏感查询的数量是AS.Count\_SQu。

**定义3 （匿名算法）**如果匿名集满足三个条件，则称为是（K，L，P）模型。条件为：（1）K-位置匿名，，其中，K= Max u.k;（2）路段多样性，，其中表示匿名域的路段数，，有L = Max u.sd；（3）P-匿名敏感信息，例如 u.p AS. Count\_SQu/。

通过应用（K，L，P）-匿名模型，可以全面保护用户在移动环境中的隐私信息。

匿名算法 AMAGS为了保护移动用户的隐私信息，MCA平台必须将移动用户向平台发送的所有查询请求分组为满足（K，L，P）-匿名模型的不同匿名集。我们提出了一种通过交换和合并用于生成匿名集的过程的算法(AMAGS).AMAGS的特征包括：

（1）构建满足k-匿名的初始匿名集;

（2）对于不满足（K，L，P）- 匿名模型的初始匿名集，交换相邻匿名集中的用户，然后合并相邻的匿名集;

（3）对于在交换和合并处理完成之后不满足（K，L，P） - 匿名模型的匿名集，添加虚拟对象使它们满足（K，L，P） - 匿名模型。

AMAGS算法的步骤如下：

（1）使用深度优先搜索算法（DFS）遍历道路网络。 在遍历中，路段和移动用户按顺序编号。

（2）选择道路段上的移动用户作为新匿名集的第一个组成部分，该移动用户未包含在任何构造的匿名集中。查找并按顺序向新匿名集中添加尚未包含在任何构造的匿名集中的相邻用户，直到该集合中的用户数量满足用户的最大匿名性需求，并且满足路段多样性的条件。

（3）重复步骤2，直到每个移动用户都包含在构建的匿名集中。

* + 1. 用户交换算法

不满足（K，L，P） - 匿名模型的初始匿名集使用交换用户的算法来处理。 处理步骤如下所述。

（1）按照不满足（K，L，P）匿名模型的匿名集中不安全用户的数量，按照升序对G\_NO1中的初始匿名集进行排序。

（2）从G\_NO1获取未处理的匿名集AS1。 将G\_NO1中未处理且与AS1相邻的匿名集提取到G\_Temp，并按照它们与AS1之间的邻接度降序排列。

（3）从AS1获取未处理的不安全用户u1。

（4）依次从G\_Temp获取匿名集AS2。

（5）从AS2获取未处理的不安全用户u2。 如果这四个条件成立：（a）u2.k  ；（b）u2.sd  AS1.Count\_S，如果u1和u2交换且AS1.Count\_S的值不会变小;（c）u2.qs  u1.qs; （d）u2.p i / K，其中i表示由AS1 中的用户发送的对u2的敏感查询的数量。

（6）如果AS1满足（K，L，P） - 匿名模型，将AS1移动到G\_OK;

重复步骤（3）到（6）通过在不同的匿名集中交换不安全的用户以优化匿名集。

（7）将G\_NO1中的剩余匿名集置于G NO2中。

在交换用户的过程中，步骤5中的四个条件确保AS2中的不安全用户u2成为AS1中的安全用户，并且AS1中的现有安全用户在交换操作之后仍然安全。也就是说，AS1中的安全用户不会因交换u1和u2而受到负面影响。如果AS1包含多个不安全用户，则按照不安全用户发送的查询的敏感度降序排序，然后获取当前未处理的不安全的用户。

|  |
| --- |
| 输入：G\_NO1,G\_OK,privacy profiles,query sensitivies  输出：G\_NO2,G\_OK  1: 按照不安全用户的数量对G\_NO1中的匿名集进行排序  2: 将G\_NO1中的所有匿名集标作未处理  3: If G\_NO1中有未处理匿名集　then  4: 取G\_NO1中未处理的匿名集AS1  5: 将AS1标记为已处理  6: 在G\_temp中排列与AS1相邻的未处理匿名集  7: for each user u1 in AS1  8: if G\_temp{} then  9: 取G\_temp 中的匿名集AS2  10: 从AS2中选择不安全用户添加到UU  11: if UU{} then  12: 从UU中选择用户u2  13: m=AS1.Count\_S //如果u1和u2交换  14： i=AS.Count\_SQU2 //如果u1和u2交换  15： if u2.k| AS1 | and u2.sdAS1.Count\_S and AS1.Count\_Sm and u2.pi/k  16: then 交换u1和u2  17: 将u2标记为安全用户  18： if AS1满足（K,L,P）匿名模型  19： then 将AS1添加到G\_OK  20： go to Step 3  21: go to Step 11  22: else go to Step 8  23: end of for statement  24: go to Step 3  26: 将G\_NO1中剩余匿名集添加到G\_NO2 |

* + 1. 用户合并算法

通过执行用户交换算法，一些初始匿名集可以满足（K,L,P）模型。G\_NO2中剩余的匿名集可以通过合并不同匿名集中的用户来进一步的优化。首先，给出p-关键用户和p-不满意用户的定义。

**定义4**（p-关键用户）对于匿名集AS中的用户u，向AS中合并一个用户ur（u.qs）后，u不再满足集敏感度需求，则称u为AS中的一个p-关键用户。

**定义5**（p-不满意用户）如果匿名集AS中的用户user不满足集敏感度需求，则这个用户为p-不满意用户。

用户的合并过程如下：

(1)将G\_NO2中的匿名集按其不安全用户的数量升序排列。

(2)将G\_NO2中的所有匿名集标作未处理。

(3)从G\_NO2获取一个未处理的匿名集AS1，提取未处理且与AS1相邻的匿名集添加到G\_Temp，按照它们和AS1之间的邻接度进行降序排列。

(4)从G\_Temp依次获得匿名集AS2。

(5)提取AS2中的不安全用户到UU。所选择的用户发送的查询请求的敏感度要小于AS1中p-关键用户和p-不满意用户的最小查询敏感度需求。

(6)从UU中取一个用户，将它合并到AS1。

(7)如果AS1满足（K,L,P）匿名模型，则将AS1添加到G\_OK。

(8)重复步骤（3）到步骤（7），通过合并不安全用户来优化匿名集。

(9)将G\_NO2的匿名集添加到G\_NO3。

|  |
| --- |
| 输入：G\_NO1,G\_OK,privacy profiles,query sensitivies  输出：G\_NO3,G\_OK  1: 按照不安全用户的数量对G\_NO2中的匿名集进行排序  2: 将G\_NO2中的所有匿名集标作未处理  3: If G\_NO2中有未处理匿名集  then  4: 取G\_NO2中未处理的匿名集AS1  5: 将AS1标记为已处理  6: 在G\_temp中排列与AS1相邻的未处理匿名集  7: if G\_temp{} then  8: 取G\_temp 中的匿名集AS2  9: m1=minimum u.qsr // u为p-关键用户  10： m2= minimum u.qsr // u为p-不满意用户  11: min\_qsr=min(min1,min2)  12: for each user u in AS2  13: if u.qsmin\_qsr  15： then 将u添加到UU  16： 从UU中选择一个用户u  17: 将u添加到AS1  18： if AS1满足（K,L,P）匿名模型  19： then 将AS1添加到G\_OK  20: go to Step 3  21: else go to Step 15  22: go to Step 7  23: 将G\_NO2中剩余匿名集添加到G\_NO3 |

1. AMAGS算法的性能实验分析

在基于真实世界道路网络环境中通过实验比较了AMAGS算法与P3RN的性能，对于先前讨论的用户的隐私信息采用三个度量：平均虚拟比率，查询成本和平均匿名时间。现有的在基于道路网络位置隐私保护工程中，P3RN似乎是解决查询敏感性问题的首选算法。我们实现的两种算法都是在JAVA中实现的。我们也在台式电脑上运行它们一个双核AMD 2.2 GHz处理器和4 GB的主存储器。

5.1 平均匿名比率

可以通过向不满足模型的匿名集添加匿名对象来增加匿名集的成功率。匿名比率是指匿名集中添加的匿名对象的百分比。 使用AMAGS算法，通过在交换和合并处理完成之后添加匿名对象，可以实现100％成功生成满足（K，L，P） - 匿名模型的匿名集。使用来计算平均匿名比率，其中num表示虚拟对象的数量，usernum表示用户的总数。

图 8a

实验结果如图8a所示. 当使用匿名比率来比较两种算法时，AMAGS的性能比P3RN差，但是随着maxk的值的增加，性能差距变小。将匿名对象添加到为P3RN设置的匿名集中是因为匿名集不满足P条件，而对于AMAGS来说可能是不满足（K，L，P）中的任何一个条件，因此需要添加更多的匿名对象。然而，与P3RN相比，AMAGS可以生成更小的匿名集和匿名域，这有助于基于匿名查询信息为LBS提供商提供更准确和有用的推荐。

5.2 查询成本

对于给定的匿名集AS，SS是其路段集合，VS是SS中的路段的端点集合。对于所有，如果存在一个端点为v且另一个端点不在VS中的道路段，则v被称为开放端点。使VS中的开放端点集合为OVS。 然后计算平均查询成本如下：（1）通过QCost（AS）= AS.Count\_S + 给定计算每个匿名性集合的查询成本，其中AS.Count\_S表示SS中的路段数量， 表示设置OVS中开放端点的数量。然后计算平均查询成本：，

其中n是匿名的数目集。

图 8b

查询成本的实验结果如图 8b所示。 两种算法的平均查询成本随着maxk的增加而增加，当maxk 10时，AMAGS的平均查询成本低于P3RN的平均查询成本。因此，当比较两种算法的查询成本时，AMAGS的性能优于P3RN的性能。 原因在于AMAGS算法对生成匿名集的过程的改进。使用P3RN，每个初始匿名集的大小是maxk，而由合并操作产生的匿名集的大小可以变为2maxk。使用AMAGS算法，初始匿名集的大小被动态地确定为不大于maxk。 合并操作将来自相邻匿名集的一个用户带到指定的匿名集中。实验结果分析表明，使用AMAGS生成的匿名集的平均大小小于使用P3RN生成的匿名集的平均大小。所以AMAGS的查询成本低于P3RN。

5.3 平均匿名时间

平均匿名时间表示算法的效率。 时间越短，算法的效率越高。 平均匿名化时间由下式给出： ，其中表示道路上的用户总数网络，T表示匿名化算法的运行时间。 图8c示出了两种匿名算法的匿名时间。 当maxk等于或大于10时，使用P3RN的匿名化时间约为0.8ms，而使用AMAGS的匿名化时间变为小于0.1ms。因此，AMAGES的效率远远高于P3RN的效率。

1. 结论

在本文中，建立了一个隐私保护服务框架，整合了多个信息服务提供商的服务资源，有助于实现参与者的双赢。可以防止移动用户的隐私信息被信息服务提供商和供应商收集和滥用，并且用户可以得到全面的信息服务。实现MCA的建议的隐私保护服务框架的关键之一是基于适当的匿名模型的匿名化过程。个性化隐私信息被定义为四元组（k，qsr，sd，p）并可用于描述移动用户的隐私要求。 根据定义的个性化隐私信息，引入并正式描述（K，L，P） - 匿名模型。

基于（K，L，P） - 匿名模型，提出了一种名为AMAGS的新的隐私保护算法，其特征在于构造初始K-匿名集，交换过程和合并过程。交换和合并的过程被逐步详细讨论和形式化的描述。 所提出的AMAGS算法用于保护移动用户的位置和标识符，但也有助于保护敏感信息。为了证明AMAGS的可行性，我们通过示例应用程序说明了每个主要步骤的结果。 使用真实的道路网络和移动用户的隐私信息，通过使用虚拟比率，查询成本和匿名化时间的度量实验分析和比较AMAGS和P3RN的性能来验证AMAGS的可行性。 实验结果表明，AMAGS查询成本和匿名化时间方面具有不同程度的优势。

**参考文献**

1. 杨璐, 田惠生, 贾明利,等. 一种高实时性的移动商务管理系统设计与实现[J]. 计算机工程, 2016, 42(1):292-299.
2. 李丹丹, 华蕊. 基于模糊综合评价法的移动商务身份认证评价[J]. 微电子学与计算机, 2016, 33(3):65-68.
3. Xue J, Liu X Y, Yang X C, et al. A Location Privacy Preserving Approach on Road Network[J]. Chinese Journal of Computers, 2011, 34(5):865-878.
4. 岑婷婷, 韩建民, 王基一,等. 隐私保护中K-匿名模型的综述[J]. 计算机工程与应用, 2008, 44(4):130-134.
5. Zhan F B, Noon C E. Shortest Path Algorithms: An Evaluation Using Real Road Networks[J]. Transportation Science, 1998, 32(1):65-73.
6. 李静, 韩建民. 一种含敏感关系社会网络隐私保护方法-(k,l)-匿名模型[J]. 小型微型计算机系统, 2013, 34(5):1003-1008.
7. 宋金玲, 赵威, 刘欣,等. k-匿名数据集的增量更新算法[J]. 计算机科学, 2010, 37(4):146-150.
8. Gangopadhyay A, Adya M. Protecting Sensitive Information in Electronic Commerce[M]// Doing Business on the Internet. 1999:77-86.
9. 陈爱东. K-匿名数据集的挖掘算法研究[D]. 东华大学, 2014.
10. Lu Q Z, Lin H L, Ge S T, et al. Wireless, remote-query, and high sensitivity Escherichia coli O157:H7 biosensor based on the recognition action of concanavalin A.[J]. Analytical Chemistry, 2009, 81(14):5846-50.
11. Hayes J, Troncoso C, Danezis G. TASP: Towards Anonymity Sets that Persist[C]// ACM. 2016.
12. 宋金玲, 刘国华, 黄立明,等. k-匿名方法中相关视图集和准标识符的求解算法[J]. 计算机研究与发展, 2009, 46(1):77-88.
13. 李晓燕, 朴春慧, 潘晓. DASS:路网上基于交换和合并的用户敏感信息保护方法[J]. 河北省科学院学报, 2014, 31(2):66-73.
14. 肖燕芳. 基于匿名区域变换的位置隐私保护模型与算法研究[D]. 华南理工大学, 2012.
15. 于娟. 数据发布中隐私保护的匿名模型及算法研究[D]. 浙江师范大学, 2010.
16. Guo K, Zhang Q. Fast clustering-based anonymization approaches with time constraints for data streams[J]. Knowledge-Based Systems, 2013, 46(1):95-108.
17. Baioco G B, Traina A J M, Traina C. An effective cost model for similarity queries in metric spaces[C]// ACM Symposium on Applied Computing. 2007:527-528.
18. Shou L, Shang X, Chen K, et al. Supporting Pattern-Preserving Anonymization for Time-Series Data[J]. IEEE Transactions on Knowledge & Data Engineering, 2013, 25(4):877-892.
19. Spinney J E. Mobile Positioning and LBS Applications[J]. Geography, 2003, 88(4):págs. 256-265.