面向 RFID 的位置隐私保护算法研究

吴婷婷,李玲娟
(南京邮电大学 计算机学院,江苏 南京 210003)

摘 要:随着移动无线技术和物联网的发展,随时随地获得个人或物品的位置信息成为可能,基于位置的隐私保护已成为当今社会中的重要问题。文中以提高隐私保护能力和位置服务效率为目标,对 RFID 追踪系统中的隐私保护问题进行研究,分析了现有的位置隐私保护方法,通过对已有算法加以改进,设计了一种高效的不依赖于可信服务器的 RFID 位置隐私保护算法。该算法用 hash 加密的方法对 ID 信息进行加密,利用垂直数据划分把时间和位置信息分别存储在不同的物理空间。理论分析和测试实验表明,所设计的算法在保护用户隐私的同时,执行效率更高。

关键词:位置隐私;隐私保护;数据库安全;RFID
中图分类号:TP311 **文献标识码**:A **文章编号**:1673-629X(2013)01-0157-04
doi:10.3969/j.issn.1673-629X.2013.01.039

Study on RFID-oriented Location Privacy Protection Algorithm

WU Ting-ting,LI Ling-juan
(College of Computer,Nanjing University of Posts and Telecommunications,Nanjing 210003,China)

Abstract:With the development of mobile wireless technology and the Internet of Things,users can achieve location information of individuals and goods in anywhere and anytime. Location-based privacy protection has become an important issue in today's society. In order to enhance the privacy protection capability and the efficiency of location-based services,it studies the privacy protection issues in RFID tracking system,analyzes the existing location privacy protection method,and designs a higher efficient location privacy protection algorithm by improving the existing algorithm. The improved algorithm is not dependent on reliable service. It encrypts the ID information with hash encryption method and stores time information and location information in separate physical space with vertical data partitioning method. The theoretical analysis and test results show that the algorithm has the higher efficiency while protecting user privacy.

Key words:location-based privacy;privacy protection;database security;RFID

0 引 言

隐私是指"与公共利益、群众利益无关的,当事人不愿他人干涉的个人私事和当事人不愿他人侵入或不便侵入的个人领域"[1]。主要包括个人身份信息、财务信息、偏好信息、家庭信息、社会信息等。随着无线设备和传感器的发展,获得用户隐私数据变得更加容易,隐私数据的保护愈发重要,其中有一个重要的方面就是用户位置隐私的保护[2]。目前,物联网产业正在兴起,而在一般的物联网应用中,采集到的数据往往具有时间和地点属性,而地点又较为重要,因此保护重要人物或物品的位置信息也是物联网隐私保护中十分关键的[3]。例如:在 RFID 供应链中,偷窃者可以通过非法攻击系统查询到重要物品目前的位置,从而达到盗窃或更换物品的目的;在利用物联网技术追踪重要人物每时每刻所在的位置以对其进行监控保护的同时,如果让不法分子查询到重要人物的具体位置,其后果是相当严重的。

国内外学者在不断的研究中提出的隐私保护算法大致可以分成两类:一类是隐藏用户的 ID 信息,使得攻击者无法识别出用户的标识信息,从而不能把用户和其所处的位置信息联系起来;另一类是保护用户的位置信息,比如用户提供给服务器的并不是其真实的位置信息,而是包含用户所在位置的一块区域[4]。在隐藏用户 ID 信息方面主要有匿名方法和使用假名等。Beresford 和 Stajano[5] 提出了 mix zone 身份保护方法,该方法将空间划分成应用区域和混合区域,在混合区域中,可以在任何时间用假名来替换自己的 ID,这样在应用区域中,用户提出请求和接受信息时可使用混合区域的假名,保护了用户隐私。位置匿名技术方面,Marco Gruteser[6] 最先运用关系数据库的数据发布隐私

收稿日期:2012-05-16;**修回日期**:2012-08-23
基金项目:国家"973"重点基础研究发展计划项目(2011CB302903)
作者简介:吴婷婷(1987-),女,硕士研究生,主要研究方向为信息安全;李玲娟,教授,博士,主要研究方向为数据挖掘、信息安全、分布式计算。

保护的 k-匿名模型,提出位置 k-匿名模型。随后许多学者也根据位置 k-匿名模型进行了算法的改进。

上述的隐私保护方法,都是基于可信数据库进行的,即原始数据已经存在于某一个可信任的数据服务器上,然后再进行数据变化及匿名。但在实际应用环境中,应该考虑没有完全可信的服务器的情况,因为随着黑客技术的快速发展,任何服务器都有可能被攻破,所以对一些极为重要的数据,要假设数据库服务器是不可信任的,需要在传感器传输数据时就对数据进行处理。另外,数据库系统本身也存在着一些安全漏洞,虽然数据库管理系统 DBMS 在操作系统 OS 的基础上增加了诸如基于权限的访问控制等安全机制,但对数据库文件本身仍缺乏有效的保护措施。黑客们仍然可以通过直接对 OS 的操作来攻击数据库文件,这类攻击很难被数据库用户察觉,因而对数据库中敏感文件进行加密处理,也是阻止这种攻击的有效手段[7]。加密方法是保护敏感数据的一个有效方法,尤其是对用户标识和一些用户敏感信息,但加密所付出的计算代价也是相当可观的。

文献[8]提出的隐私保护算法与上述研究思想不同,它是一种基于不可信数据库的算法,通过加密和分割数据库中的元组并分别存储来达到保护用户位置隐私的目的。但是,该算法在对加密数据库进行查询时要付出很大的代价[9]。

由于传感器和 RFID 的计算能力有限,寻找一种既节约计算开销又有良好保密性能的加密算法显得十分必要。

为此,文中以进一步提高隐私保护算法执行效率为目标,对文献[8]提出的算法加以改进,设计了一种新的基于数据分割和加密的 RFID 位置隐私保护算法(DPERLPP 算法)。

1 算法设计

1.1 问题的提出

T. Rodden, A. Friday[10] 等人提出了对重要的数据元组进行划分并分别存储在不同服务器上的数据库划分思想。而 Chiu C. Tan[8] 等人运用这种思想提出了一种基于 RFID 追踪系统的保护方法,主要研究已经存储于数据库中的数据的隐私保护。在 RFID 物体追踪系统中,物体标签往往携带着关于物体的重要信息,包括标签 ID 号、标签自带的密钥等,RFID 标签也应具有简单的数字生成器、加和乘的功能。Chiu C. Tan 等人针对的是一个不依赖于可信任服务器的 RFID 追踪系统,在 RFID 追踪系统中构建两张数据表:时间表 TS(time, n)和位置表 LS(W, location)。TS 表用于存储标签的时间信息,两个属性分别表示时间和随机数。LS 用于存储标签的路径信息,两个属性分别表示 ID 标志和位置信息。两张表分别存储在不同的物理空间。

如果用户想要查询自己标签在 10 点钟所在的位置,首先要找到 10 点时的 n,这就要通过查询时间表 TS 得到,查询过程如下:

1)首先根据已知的 ct 生成 $n=h(s,ct)$,s 是标签自带的密钥,ct 是简单的计数器生成的数,每读一次标签,计数器增加 1,函数 $h()$ 是 hash 函数。根据生成的 n 查找 TS,利用二分查找法,如得到的时间比 10 点大,则减小 ct 的大小,重新计算 n;如比 10 点小,则增大 ct 再重新计算 n,然后继续查找 TS,反复查找两张表,直到返回的值正好等于 10 点,取此时的 n;

2)利用此 n 计算 h(ID, n)得到 W,W 是加密过的用户标记,再利用 W 查询 LS 表得到 10 点时标签所在的位置。用户在进行二分搜索时会生成一张用户表(见表 1)。

表 1 用户查询时生成的表

ct	n	time
…	…	…
ct_i	$h(s, ct_i)$	?
ct_{i+1}	$h(s,ct_{i+1})$	?
…	…	…

这种查询方法使攻击者很难获取到标签的位置信息,达到了隐私保护的目的,但查找 n 的代价很大,尤其当 TS 表很大时,可能生成很大的用户表。加上 TS 和 LS 分别存储于不同的物理空间,查询所消耗的通信代价也很可观,这在物联网巨大的数据量及无线网络有限的计算资源面前是一个不容忽视的问题。

针对以上问题,文中设计了一种新的基于数据分割和加密的 RFID 位置隐私保护算法(Data Partition and Encryption Based RFID Location Privacy Protection Algorithm, DPERLPP)。

1.2 DPERLPP 算法描述

假设 RFID 标签上有自带密钥 $k1$、$k2$,自动计数器的值用 ct 表示。时间表 TS 加入一个标志字段 W1,TS 中存放标志 W1、每一次读取标签的时间 Time 和随机数 N,这个 N 是标签中计数器值 ct 和标签自带密钥 $k1$ 经过 hash($k1$,ct)产生的,W1 是由标签的 ID 和 n 经过 hash(ID, n)产生的。LS 中存放随机的 W2 和标签被读取的地点 Location,W2 是由标签自带密钥 $k2$ 和 n 经过 hash(n,$k2$)产生的。改进后生成的时间表和位置表的结构见表 2 和表 3。

hash 函数具有如下特性:不可从消息摘要中复原信息;两个不同的消息不会产生同样的消息摘要[11]。利用 hash 函数的这种特征,为 LS 和 TS 数据库中的 W1、W2 和 n 进行单向加密,因为无需解密,所以利用

表 2　改进的时间表 TS

W1	Time	*N*
$h(ID, ni-1)$	10:00 am	$h(k1, cti-1)$
$h(ID, ni)$	10:00 am	$h(k1, cti)$
$h(ID, ni+1)$	10:15 am	$h(k1, cti+1)$
…	…	…

表 3　改进的位置表 LS

W2	Location
$h(ni-1, k2)$	Office 1
$h(ni, k2)$	Office 3
$h(ni+1, k2)$	Office 2

hash 加密函数是比较安全和快捷的方法。同时,位置表 LS 中的位置标识属性 W2 经过两次 hash 的保护,即使攻击者们获得数据库的信息,也很难推测出 W2 所对应的 RFID 标签的 ID 及其所在的位置。

查询 10 点钟标签所在的位置,只需用下面的语句得到对应的 ni:

Select * from TS where $N = ni$ and Time = 10:00am, where hash(ID, ni) = W1;

然后再用下面的语句查询地点:

Select * from LS where W2 = hash($k2$, ni)

查询过程如下:

(1)首先根据已知的 ct 生成 $n = h(k1, ct)$,再用 ID 信息和 n 进行 hash 运算得到 W1,通过 W1、time、n 这三个已知量进行查找,得到 10 点钟对应的 n 值;

(2)利用得到的 n 值在 LS 表中查询标签在此时所对应的位置信息。

这样在用户查询过程中就不必产生用户表,省去了生成用户表的时间。

2　测试实验

为了测试文中提出的 DPERLPP 算法的效率,文中设计了相关实验。实验使用了 Windows 7 操作系统、MYSQL 数据库管理系统、Java 语言。

数据量从 5000 到 50000 条,均由系统经过 hash 加密随机生成。数据表 TS 和 LS 存储在不同物理空间时,在同样查询下产生的通信代价是相同的,所以本实验只考虑单机运行的查询时间。

实验对最好情况(数据在数据库中的存储最利于二分搜索的情况)、最坏情况(数据在数据库中的存储最不利于二分搜索的情况)、平均情况(以上两种情况的折中)下文献[8]的算法和文中提出的 DPERLPP 算法的执行效率进行了比较。Test1 表示文献[8]的算法的查询时间,test2 表示 DPERLPP 算法的查询时间。

图 1 是两种算法在最好情况下查询时间的比较结果。可以看出,随着数据量变大,test1 和 test2 在最好情况下查询时间相差不大,趋势都相对平稳。

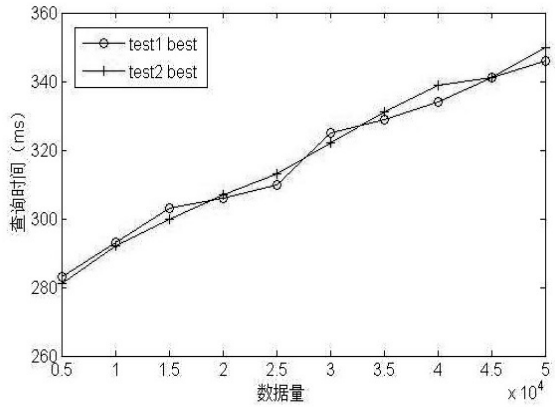

图 1　两种算法最好情况下的查询时间

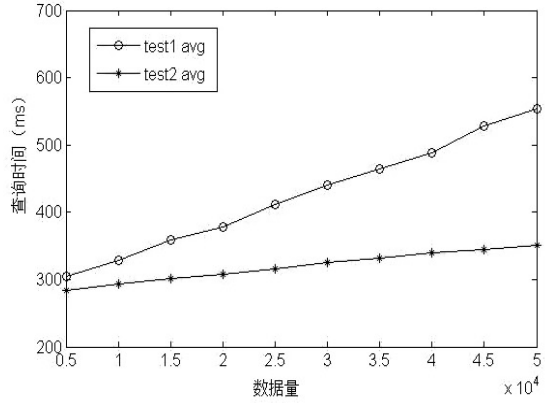

图 2　两种算法平均情况下的查询时间

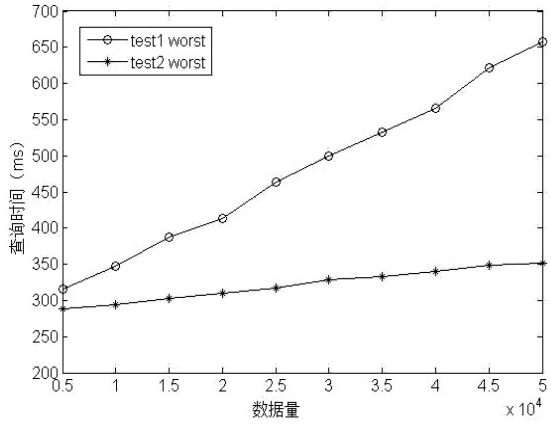

图 3　两种算法最坏情况下的查询时间

图 2 表明了在平均情况下,随着数据量的增大,test1 呈递增趋势,而 test2 相对平稳。

图 3 表明了在最坏情况下,随着数据量的增大,test1 不断递增,而 test2 相对平稳。

此外,从实验结果可以看出,test2 在三种情况下的查询时间相对平稳,都与文献[8]的算法最好情况下的查询时间相似。

3　算法分析

3.1　算法时效性分析

由图 1 展示的实验结果可见,DPERLPP 算法与文

献[8]的算法相比,在时间效率上有显著的提高,这是由于 DPERLPP 算法中通过加入标识 W1 简化了查询过程,使得用户在查询时不需使用二分搜索来产生用户表,这一步是查询中节省时间的关键。文献[8]的算法在最坏情况下,用户进行二分搜索时会耗费大量的时间,因而在大数据量情况下,DPERLPP 算法会比文献[8]的算法节约大量的执行时间。

3.2 算法隐私保护性分析

就隐私保护度的角度来看:

(1)存储时间表 TS 的数据服务器遭到攻击时,由于 W1 经过 hash 加密,并且每次加密的 n 都不同,即使是同一 ID 在不同时间加密后结果也不同,所以攻击者并不知道某一个标识 W1 对应的真实 ID 信息,也不知道哪两个 W1 是对应同一个 ID 标识的,因此当存储时间表 TS 的服务器被攻击后,攻击者无法查询出 RFID 标签在某一时刻对应的 n 值和这一时刻的位置信息,更没法查询标签的运动轨迹。在特殊情况下,假设攻击者已经知道足够多的背景知识,知道了哪几个 W1 的值对应的是同一标签,但如果攻击者不知道这一 RFID 的密钥 $k2$,欲查询位置信息时,还是推测不出 LS 表中 W2 的值,得不到与其对应的位置信息。

(2)存储位置信息 LS 表的数据库服务器遭到攻击时,由于 W2 是 n 和 $k2$ 的 hash 运算得到, n 是由计数器 ct 和用户密钥 $k1$ 经 hash 运算得到的,而攻击者并不知道用户的密钥 $k1$ 和 $k2$,计数器的值也要经过大量运算才能推测到,所以理论上是无法得到标签与位置的对应信息的。这样,就把用户的标识信息完全加密隐藏,即使攻击者获得数据库里的数据,也无法得到自己想要的信息。

(3)如果攻击者同时攻击时间服务器和位置服务器时,可以得到的信息有某一标签在某一时刻的 n,如果想知道这一标签的位置信息,还要知道它的密钥 $k2$,但攻击者无法获知 $k2$ 信息,也无法从已知数据中推测出来。特殊情况下,当攻击者通过观察连续的查询消息和获知一定的背景知识时,可以关联出哪些标识对应的是同一 RFID 信息,这样就可以通过已知的 n 通过大量运算推测出 $k2$,从而就能查询到物品的位置信息。假设以上所有特殊情况都成立的情况下,攻击者也需要花费大量的时间来获得背景知识和破解加密的用户信息,其实这对攻击者来说并不容易实现。实际应用中,由于时间表和位置表存储在不同的物理空间,这给攻击者同时攻击两个数据库也带来了难度。

由上述分析可得,文中的 DPERLPP 算法在保护位置隐私的同时,时间效率上有显著的提高。

4 结束语

文中通过对已有的 RFID 追踪系统的隐私保护算法加以改进,设计了一种高效的基于数据分割和加密的 RFID 位置隐私保护算法,理论分析和实验结果表明,该算法在具有良好的隐私保护性的同时也具有较低的时间复杂度。

但是,有必要进一步设计针对攻击者同时攻击时间和位置两个数据库时的位置隐私保护措施,比如进一步对用户位置信息进行模糊化或匿名[12]等处理。事实上,随着物联网产业的发展,位置隐私保护的研究将具有很大的空间和很好的应用前景。

参考文献:

[1] 王利明. 人格权法新论[M]. 长春:吉林人民出版社,1994.

[2] 潘 晓,肖 珍,孟小峰. 位置隐私研究综述[J]. 计算机科学与探索,2007,1(3):268-281.

[3] Sen J. Privacy Preservation Technologies in Internet of Things [J]. Journal BITM Transactions on EECC,2009,1(4):496-504.

[4] 魏 琼,卢炎生. 位置隐私保护技术研究进展[J]. 计算机科学,2008,35(9):21-25.

[5] Beresford A R,Stajano F. Location privacy in pervasive computing[J]. IEEE Pervasive Computing,2003,2(1):46-55.

[6] Gruteser M,Grunwald D. Anonymous Usage of Location-based and Services through Spatial and Temporal Cloaking[C]// Proc. of the 1st International Conference on Mobile Systems, Applications and Services. New York:ACM,2003:163-168.

[7] 杨 勇,方 勇,周安民. 秘密同态技术研究及其算法实现[J]. 计算机工程,2005,32(2):157-159.

[8] Tan C C,Xie Lei,Li Qun. Privacy Protection for RFID-based Tracking Systems[C]//IEEE International Conference on RFID. [s. l.]:[s. n.],2010:53-60.

[9] Yang Z,Zhong S,Deng R H,et al. Privacy-preserving queries on encrypted data[C]//European Symposium on Research in Computer Security (ESORICS). [s. l.]:[s. n.],2006.

[10] Rodden T,Friday A,Muller H,et al. A lightweight approach to managing privacy in location-based services[J/OL]. 2002. http://comp. eprints. lancs. ac. uk/1475/1/rodden - lightweightprivacy-2002. pdf.

[11] 李 哲,方 勇,陈淑敏,等. 数据库加密技术中散列函数的应用[J]. 计算机工程,2003,29(17):68-70.

[12] 徐正峰,杨 庚. LBS 中基于标识符的连续查询模型研究[J]. 计算机技术与发展,2011,21(9):237-241.

面向 RFID 的位置隐私保护算法研究

作者： 吴婷婷， 李玲娟
作者单位： 南京邮电大学 计算机学院,江苏 南京 210003
刊名： 计算机技术与发展
英文刊名： Computer Technology and Development
年，卷(期)： 2013(1)