DOI:10.13878/j.cnki.jnuist.2017.05.014

张钊[1]　华景煜[1]

基于指纹识别的室内定位中的隐私保护

摘要

基于指纹识别的定位是最流行的室内定位方法.在离线阶段,服务器测量指纹,比如来自特定空间已知位置的不同接入点(AP)的接收信号强度(RSS),测量后服务器将测量结果保存在数据库中;在线上阶段,用户同时向服务器发送他当前指纹的测量结果以及位置查询请求,服务器将在数据库中查找与测量结果最接近的指纹.虽然这种方法已经被研究了很久,但现有的工作并没有考虑2个隐私要求:供应商希望保护他们花大代价收集的指纹,用户想要对服务器保留他们的指纹测量结果,以避免泄漏位置.为了实现隐私保护,本文提出一种使用加密技术的指纹匹配方案,这个方案在加密情况下计算由用户测量的指纹与服务器存储的指纹的距离,服务器存储的指纹在这一过程中仍处于密文空间.本文证明了这个方案在进行单点定位时能够很好地保证两者的隐私要求.为了减少高昂的时间开销,本文还提出了一个基于网格划分的改进方案,以及以有限的隐私损失为代价的扩展方案.为加强安全性,最后提出了有效对抗特定攻击的对策,在这种攻击中恶意用户可以通过重复定位获得服务器存储的指纹.使用公众RSS指纹数据集的扩展实验结果显示本文方案足以在实现实时定位的同时保留定位精度.

关键词

基于指纹定位;室内定位;隐私保护

中图分类号 O429

文献标志码 A

收稿日期 2017-07-20

资助项目 国家自然科学基金(11471003,61425024)

作者简介

张钊,男,硕士生,研究方向为移动安全与隐私保护.ericzz0727@gmail.com

华景煜(通信作者),男,博士,助理研究员,目前研究方向为移动安全、隐私保护等.huajingyu@nju.edu.cn

1　南京大学　计算机科学与技术系,南京,210023

0　引言

室内定位对于众多基于位置的手机应用来说非常关键.在现有的室内定位方法中,Wi-Fi信号指纹由于其高精度最受关注[1].这种方法通常由2个阶段组成:训练阶段和使用阶段.在训练阶段,服务提供商测量指纹,即来自多个接入点(APs)的Wi-Fi信号强度,在空间中进行多个位置采样并将其存储在数据库中.指纹也可能包括其他环境特征,如采样点的声音、光线、颜色等[2-3].在使用阶段,当用户定位自己的时候,同样采集关于他现在位置的指纹,然后要求数据库中最匹配的指纹,并根据返回的指纹来估算他的位置.由于指纹数据库不容易构建,服务者通常把他们作为机密,出于商业利益考量不愿意将这些数据公布于众,因此他们通常将指纹数据库放在他们完全控制的中央服务器上.用户被要求将测量数据上传到服务器而服务商会返回一个最接近的数据给他们,用户没有任何权限来直接访问数据库.然而这种方法虽然保护了服务商的利益,但是由于服务商可以完全跟踪用户的位置轨迹,用户会担忧自己的隐私问题.因此理想的基于指纹识别的方案应该同时保护服务提供商和用户的隐私.这个任务具有挑战性,因为用户隐私和服务商的隐私在大多数时间内是冲突的,据我们所知目前的方案都没有考虑这个问题.在本文中,我们试图提出一种基于加密的高效率的保密方案来填补这方面的空白.主要工作如下:

首先提出了一种基础的保护隐私的指纹匹配方案.这种方案使用ElGamal加密方案[4]经过全同态方案来计算用户测量的指纹和服务器存储的指纹之间的距离,这一过程中服务器的指纹仍然是加密的.

我们随后修改了基本方案,通过损失一点隐私的代价减少时间消耗.我们将原始空间在地理位置上划分为 n 个大小类似的网格,记录每个网格的中心指纹.用户想要找到自己的时候先通过秘密比对他的指纹和每个中心的指纹来决定他属于哪个网格.由于这个方案不需要与服务器的所有指纹比较,因此时间开销显著降低.此外,因为服务器只知道用户属于这 k 个区域的某一个,因此只要用户不会频繁发起请求,用户只会丢失有限的隐私.

但是上述2个假设都难以在现实世界中得到保证,所以我们提出进一步的方案来改进这些冲突.使用聚类技术自动分类指纹来代替手工划分网格,可以大大降低定位错误.如果一个用户经常执行定位,并

始终使用随机策略选择虚拟网格,恶意服务商可以将他运动的一些特征以及选取的多个定位相关联,在短期内确定他真正的所处的网格.我们提出有效的对策来抵御这种关联攻击.基础方案和改进方案都需要服务商将数据库中的指纹对应位置公开,这可能仍会损害服务商的隐私,因此我们设计了第3个扩展方案来解决这个问题.之后,研究了重复定位攻击,并且提出了一个针对基于聚类的设计方案的改进方案,可使恶意用户很难通过少量的定位来获得服务器的指纹数据.

最后,本文进行了大量的实验来评估方案的表现,使用了大概1 000个真实的公共数据集,展示了改进方案以及基于聚类的改进方案都可以在1 s内完成一次定位,并且指纹集的提升对于时间开销提高不大,因此对于更大的空间具有良好的扩展性.结果显示只要在集合中引入一些重叠,基于聚类的扩展可以正确找到95%以上的最接近指纹.我们还实际测量了对重复定位攻击的影响,实验结果显示了本文方案显著增加了攻击者获得服务器指纹的难度,而付出的定位精度损失可以忽略不计.

本文的其他部分安排如下:第1部分描述相关工作;第2部分给出了基于指纹定位的隐私问题的正式定义;第3部分展示了使用同态加密的基础隐私方案;第4部分提出了权衡效率和隐私的改进方案;第5部分描述了3个扩展;第6部分讲述了重复定位攻击;第7部分展示了实验结果以及评估;第8部分进行了总结.

1 相关工作

定位相关的隐私保护在近几年的研究中成果非常多,然而大多数工作都集中在基于位置的服务(LBS)中的用户位置保护上,而对于用户定位过程没有研究.在这部分我们会总结一下目前在基于位置服务(LBS)中的位置隐私保护机制(LPPMs),这可能会给我们设计保护用户位置的定位系统一些启发.

在基于位置服务中,用户被要求将他的位置提交给服务商,服务商会基于位置提供一些后续服务,比如说常用的地图服务.基于位置服务中的位置隐私保护机制(LPPMs)的目标是在使用用户能够使用这些服务的同时尽可能少地暴露位置信息[5],这与我们允许用户获取服务器的最接近指纹的同时保护用户自身指纹信息的目标很相似.

现在最流行的LPPMs工作是在用户上传位置信息到服务器之前进行混淆[6-9],这样会导致定位精度有损失,这在对精度要求更高的室内定位中可能难以接受.在另一类的LPPMs工作中通过混淆区域来隐藏用户的位置[10-11],此种方案的资源开销非常高昂.还有一种类型的LPPMs工作是在实际服务请求发起的同时发起几个虚假请求来干扰恶意服务提供商[12],这种方法的局限之处在于当用户对于数据库中的指纹结构不能完全了解的情况下,伪造的指纹很容易被攻击者过滤掉虚假请求.最后一种类型的LPPMs采用加密技术隐藏用户隐私[13-15],该种机制的关键思想是利用一些同态加密来将涉及位置的计算变为密文空间的计算,从而不泄露原始值.本文使用最后一种方法来实现指纹定位中的隐私保护,其最大的挑战是怎样设计一种方法实现模糊指纹从明文空间到密文空间的映射.此外,很多高级匹配算法在密文空间中都无法使用,因此时间开销通常非常高.

在5.2中,我们使用聚类技术来自动对指纹进行分组以此减少需要匹配的指纹数量,这种方法最初由Swangmuang等[16]提出.他们使用这种技术来减少基于指纹的室内定位分析模型的计算量.Altintas等[17]使用聚类来改进定位漏洞,但是都没有涉及本文在5.2中提出的问题.

2 问题陈述

如前所述,在基于指纹识别的定位中隐私问题包括2个方面:服务器想要阻止用户学习他收集到的指纹,而用户想要防止服务器知道他测量的指纹,进而推测出他的位置.我们假设每个指纹表示为一个包含 n 个特征的向量,假如一个区域内有 n 个AP,那么每个特征就可以代表着一个区域内的一个特定AP的信号强度.我们想要解决的问题可以表述如下:

给定一个在服务器 S 上的指纹数据库 $D(a_1,a_2,\cdots,a_m)$,一个由用户 U 测量到的指纹 a_u.我们设计一个在 S 和 U 之间的定位协议,在这个协议中 S 和 U 分别将 D 和 a_u 作为输入.在使用这个协议后,U 可以获得 $loc(a_{\text{closest}})$ 作为 D 中距离 a_u 最近的指纹对应的位置.对于任意 i,$1 \leqslant i \leqslant m$,$Dist(a_u,a_{\text{closest}}) \leqslant Dist(a_u,a_i)$.

在这里,$Dist$ 是表示2个指纹的距离的函数,本文直接使用欧几里得距离 $Dist(a_u,a_i)=\sum_{t=1}^{n}(a_u \mid t$

$|-a_i|t|)^2$.在这个过程中,协议必须保证2点:1)U不能获得可以帮他获取D中任何一个a_i的任何信息;2)S不能获得任何能帮他推测a_u或者$Dist(a_u,a_i)$值的信息.

我们将从第3部分开始解决上述问题.特别的,我们首先会呈现一个通过同态加密实现的基础版本,可以很好地满足用户和服务商的隐私需求.这个版本因为高时间消耗而不能应用在大规模的室内定位中,因此我们又进一步呈现了更加适用的版本,通过消耗少量的隐私来减少时间消耗.

3 使用同态加密的基础隐私保护方案

首先描述一下基础隐私保护方案.根据上文的描述,基于指纹定位的主要目标是获得服务器数据库中与用户测量指纹最接近的指纹,即计算用户获取的指纹与服务器数据库中指纹的距离.本文的基础方案利用ElGamal的同态加密方案[4]在密文空间计算距离,这个过程中服务器和用户都无法获得对方的指纹.

图1大致说明了基础方案.根据ElGamal的方案,一个用户在定位自己的时候需要先生成一对密钥(k_s,k_p),其中k_s作为私钥,而$k_p=(p,g,h)$作为公钥.同态加密具有如下性质:$E_{k_p}(m_1\cdot m_2)=E_{k_p}(m_1)\cdot E_{k_p}(m_2)$.用户将公钥发送给服务器,对于每个用户测到的指纹$a_u$中的特征$a_u[t]$,用户通过公钥计算$e_u[t]=E_{k_p}(g^{a_u[t]})$和$e_u'[t]=E_{k_p}(g^{a_u[t]})^2$,并且将这2个值发送给服务器,服务器对于数据库中的每个指纹a_i进行如下计算:

$$c_i=\prod_{t=1}^{n}e_u'[t]E_{k_p}(g^{(a_i[t])^2})e_u[t]^{-2a_i[t]}. \tag{1}$$

根据ElGamal的同态加密的性质,我们可以推导出$c_i=E_{k_p}(g^{Dist(a_u,a_i)})$,服务器返回所有的结果$C=\{c_1,c_2,\cdots,c_m\}$给用户,用户可以解密这些结果获得$D_{k_s}(C)=\{g^{Dist(a_u,a_1)},g^{Dist(a_u,a_2)},\cdots,g^{Dist(a_u,a_m)}\}$.

我们认为a_u和最近指纹a_{closest}的距离应该足够小以保证离线阶段的数据密度.用户对$D_{k_s}(c_i)$与集合$\{g^1,g^2,\cdots,g^k\}$的值进行比较,当相等时表示$Dist(a_u,a_i)$与l相等,如果没有找到相等的数则将$Dist(a_u,a_i)$设为无穷大.当获得所有距离的时候用户就知道数据库中的哪个指纹和他最接近.这里我们假设用户可以获得指纹集合对应的位置集合,但是无法知道每个指纹的特定位置.我们在5.3部分会考虑服务商将上述信息作为隐私的情况下保护这些位置.

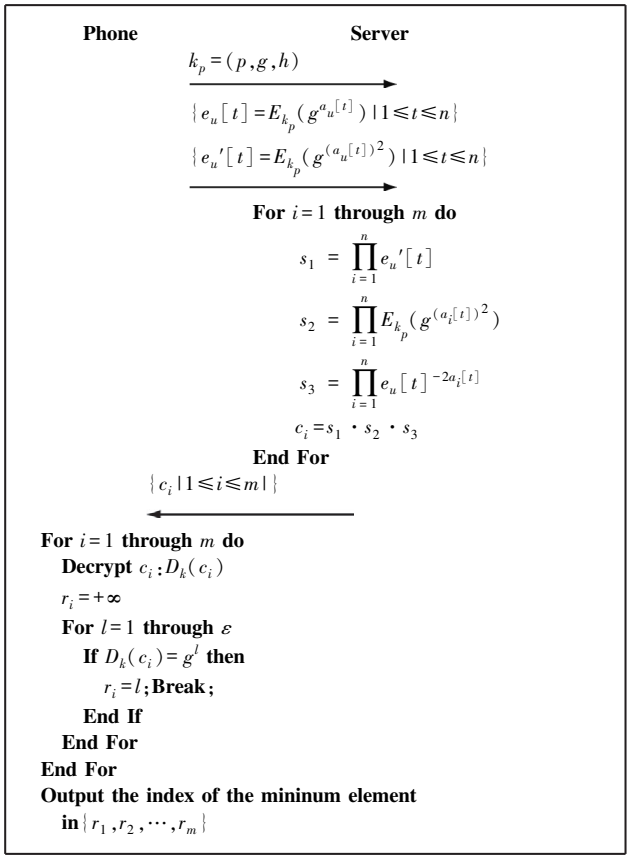

图1 基础方案流程
Fig. 1 Overview of the basic scheme

4 划分网格的改进方案

基础方案在应用到大范围的空间中时会有很大的计算消耗,接下来使用改进的高效率方案,工作的关键在于减少需要进行比对的指纹数量.在这里假设目标空间有少量障碍物,在下部分还会减弱这种假设条件.

4.1 方案描述

在新方案中,服务器在对用户提供服务前会首先将空间划分为w个大小相同的网格,用$G_1,G_2,\cdots,G_w$表示,必须保证每个网格包含相同数量的指纹.在每个网格中,我们都会找出距离中心最近的指纹,并将这w个最近的指纹为中心指纹,表示为$P=\{p_1,p_2,\cdots,p_w\}$.图2讲述了定位步骤,主要分为如下4步:

第1步:用户使用图1中描述的基础方案获得中心指纹p_α.因为我们假设空间有少量障碍物,因此更近的指纹距离通常意味着更近的地理位置.因此可以推测D中最近的指纹很大可能和p_α在一个网

格中.

第2步:用户随机选择P中的其他$k-1$个指纹,将这$k-1$个指纹和p_α一起发送给服务器.注意用户其实并不知道这些指纹的具体值,因此只是发送他们的下标.服务器可以知道其中有一个是p_{closest},但是无法获得p_α的值.

第3步:服务器计算收到的k个中心指纹的距离,我们使用C_x分别表示G_x个指纹的结果,这些结果一起返回给用户.

第4步:用户解密C中的所有元素,然后获得a_u到各个中心指纹的距离,从其中选择最短距离的指纹作为结果,丢弃其他的返回值.

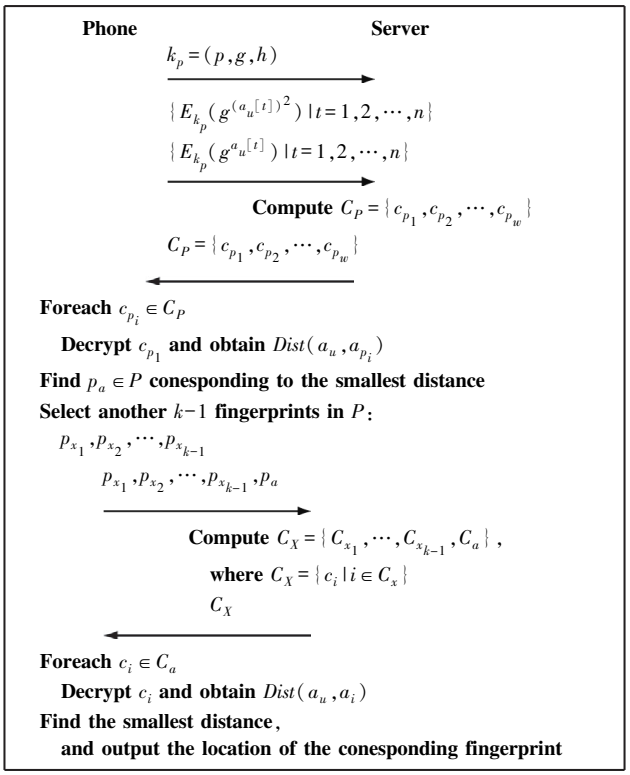

图2 改进方案流程

Fig. 2 Overview of the practical scheme

在这个方案中,由于不需要对所有指纹进行加密运算,因此时间效率会有很显著的提高.但由于服务器可以知道用户的位置就在k个网格中,所以牺牲了一定的用户隐私.因此这个方案实际上对于用户隐私和效率进行了一定的折中,这个折中主要由k和w来控制.当k值比较大或者w值比较小的时候隐私性会更好.

如果我们假设m个指纹会独立分布在w个网格中,那么此方案中服务器需要的运算量很容易估计,是$mk/2w$.

4.2 隐私分析

现在更详细分析一下这个方案的隐私开销.

对于服务器来说,除了收到与基础方案相同的信息以外,还收到了额外的k个中心指纹,即k个网格,目标手机就在其中的一个网格上.由于服务器并不知道哪一个网格包含了目标手机,如果k值不是特别小,那么用户隐私泄露的可能性就会很小.注意这个结论只有当用户不会频繁请求定位时才会成立.如果2个连续的定位请求在短时间内发起,服务器仍然有可能获得准确的位置.下一部分我们还会讨论这个问题的细节.事实上即使服务器成功预测了包含用户的网格,当网格比较大时服务器还是只能预测用户的大概位置.

对于用户来说,他所有能收到的信息都是基础方案上的一个子集.因此,如果基础方案里他无法获得服务器的指纹特征,在这里用户还是无法获得指纹特征.

5 更加现实的改进方案

前文提到的改进方案依赖一些假设,这些假设在现实世界中未必成立.因此我们提出一些改进方案来弱化这些假设.

5.1 改进网格选择策略

在上述改进方案的第2步中,随机选择$k-1$个中心指纹来进行混淆,但是这些随机选择的指纹可能会有2个问题:

第1个问题是不能保证选择的这$k-1$个网格是均匀分布的,他们和用户所在的网格可能恰好相邻,因此恶意服务商可以将用户锁定在一个小的范围内.我们可以均匀划分原始空间为k个非重叠的区域,然后让用户随机选择每个区域中的一个网格作为虚拟网格.

第2个问题是在这种策略下如果用户在进行导航,持续对自己进行定位,那么恶意服务商可以通过用户轨迹来确定真实位置.我们将服务商的这种攻击称为相关攻击.我们针对相关攻击制定了防御策略,基本思想是模拟出$k-1$条人的行走轨迹,根据这些模拟轨迹提出定位请求,这使得恶意服务商定位用户的难度大大增加.

5.2 基于聚类的扩展

前文的改进方案都是在不同目标空间存在少量障碍物以及指纹距离相近对应物理位置相近的假设上的,然而现实的室内环境通常会有很多障碍物,这

会影响距离和信号强度之间的关系.服务器可以使用一些经典的自动聚类方法来完成指纹的自动划分,这样可以保证划分出来的网格内的指纹会具有更好的相似性.此时当一个用户测量的指纹发现属于某个分组时,与他相近的指纹有很大的可能属于同一分组.更重要的是这种方法不需要对指纹距离和物理空间距离的关系做出任何假设,而且不需要消耗任何精力来考虑墙体以及地板等环境影响,所以这种方法会更加高效精确.

RSS 指纹聚类[16-17]被用来改进传统基于指纹定位的适用性,我们可以直接使用现有的方案,大多数方案都由 k-means 方法改进而来.在这里只提出 2 个前面没有被提及的重要问题:

第 1 个问题是考虑定位的时间效率问题,我们希望每个分组会有同样多数量的指纹,然而 k-means 聚类方法并不能保证这一点.k-means 算法可能会产生不平衡的分组,可以通过递归调用聚类算法来解决这个问题.如果有分组的元素数量大于设定的阈值 $mk/2w$ 时,将对这个分组再次使用聚类算法进行分组.

第 2 个问题是一个用户测量到的指纹还可能会和他在数据库中最接近的指纹处于不同的分组中,这会导致最后的定位产生很大的偏差.我们的解决方法是考虑所有与目标指纹接近的分组,这样不管目标指纹被划分到哪一组都可以找到距离最近的指纹.不过这样需要计算更多的指纹,会导致时间效率下降.

5.3 保护数据库指纹的信息

上述的所有方案都假设数据库中的指纹对应的位置是公开的.但是当一个用户知道距离他最近的指纹时,他很大程度上可以估计出这个指纹对应的位置,有些服务商并不乐意看到这种结果.因此需要一个方案使得用户可以根据最近的指纹获取位置的同时无法获取服务商指纹的其他信息.

我们仍然使用同态加密方案来解决这个问题.假设用户根据前文方案知道距离自己最近的指纹编号是 i,他想要得到这个指纹的位置.我们的隐私保护策略如下:

1)用户计算 $R=E_{k_p}(g^{-i})$ 并把值发送给服务商;

2)服务商收到 R 以后使用一个随机指数 s,计算 $R_j'=(R\cdot E_{k_p}(g))^s$,其中 j 取值为 1 到 m,然后对于每个数据库中的指纹 a_j 计算 $R_j''=R_j'E_{k_p}(loc(a_j))$ 并发送给用户;

3)用户可以通过解密 R_j'' 来获得距离他最近的指纹位置.

6 防御重复定位攻击

用户可能通过不断给服务商发送重复定位请求来获取一些服务商的指纹信息,我们在这一部分将针对这种攻击提出防御方案.当使用聚类扩展方案后,服务商的指纹会被分成不同的组.如上文提到的每个用户手机在一个地方实际上只能检测到少量的 AP 信号,指纹信息所构成的向量是稀疏的.我们用 S_{G_i} 表示在 G_i 分组中至少有一个指纹存在非零值的特征的集合.因为在同一个组内的指纹通常比较接近,因此 S_{G_i} 的大小通常会远小于所有的特征数量,即所有空间区域的 AP 数量.我们可以仅使用 S_{G_i} 中的特征来从 G_i 中找到距离用户测量结果最接近的指纹.利用这个结论来抵御重复定位攻击,这需要付出一些定位精度的代价.

特别的,当服务器计算用户测量结果和 G_i 集合中的指纹的距离时,我们从 S_{G_i} 中随机选择 p 的特征来进行计算.因为用来计算的特征是服务器随机选择的,攻击者无法保证 2 个结果使用一个特征集合算出来,因此他们就无法根据定位结果推测出向量的特征值.如果假设 S_{G_i} 的大小是 k,那么服务器在 2 次定位中选择同一个属性集合的概率是 $1/C(k,kp)$,这概率在现实中非常小.因此可以认为这个措施可以大幅增加攻击者计算服务器指纹的难度,在后文实验中会提到这一点.

因为我们选择的是用部分特征来计算距离,这一定会影响到最后的定位精度.根据后文的实验结果,只要选取的 p 值不太小,精度影响很有限.

7 结果评估

在这一部分使用真实的 RSS 指纹数据集来评估本文基础方案和改进方案的表现.我们使用 JCE 框架[18]和 Bouncy Castle 库(http://www.bouncycastle.org/)来实现 ElGamal 方案,并使用了 ICDM'07 DMC[19]的公开数据集来作为我们的指纹数据库内容.这个数据集包含了 1 400 个带位置标签的指纹,这些指纹都由同一个设备在同一建筑的 200 个不同地点采集.每个位置被划分为 1.5 m×1.5 m 的网格,每个位置都采集了几个不同的指纹.我们使用了 256 bit 长度的密钥.

7.1 时间效率

首先计算基础方案、划分网格的改进方案以及基于聚类的改进方案的时间效率.聚类的改进方案中,将所有的指纹聚类结果划分为 15 个组,每一分组的指纹数量都在 40~70 之间,每 2 个分组之间都有大概 10 个元素的交集.我们在 2 个实验中都选择了 4 个混淆网格或者分组.用户指纹都是从数据集中随机提取的,所有的数据都是经过 10 次实验得出的结果.

分别计算离线阶段和使用阶段服务商和用户的时间开销.离线阶段 3 个方法的开销都一致,经过实验统计,离线阶段用户的时间消耗大概为 1 s,服务器端的时间消耗大概为 16 s.因为这是在使用之前部署的,所以这些时间消耗并不重要.图 3 给出了使用阶段的时间开销,可见改进方案显著提高了时间效率,用户可以在 1 s 以内使用 2 个方案获得他的位置,而基础方案需要大约 2 s 的时间.

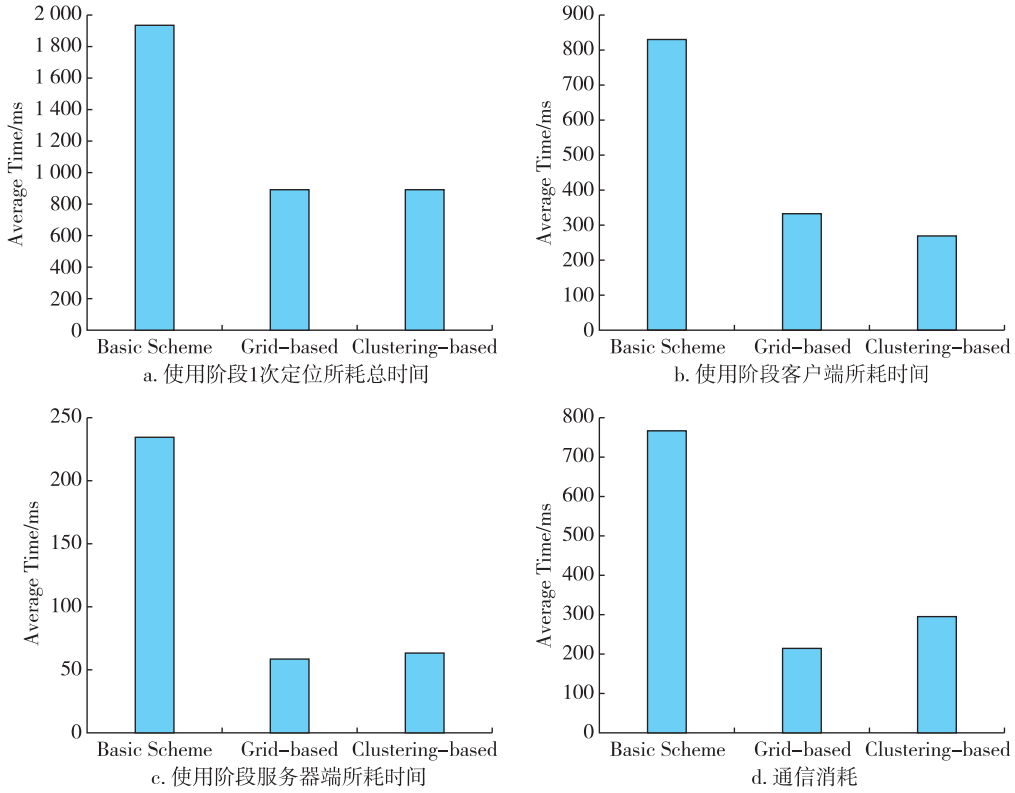

图 3 3 个方案使用阶段时间消耗

Fig. 3 Online time costs of the three proposed schemes

为了研究指纹数据库对时间的影响,将数据集从 500 逐步增加到 800,图 4 给出了新的时间消耗.图 4 显示这一时间的增加是线性的,这与理论一致.

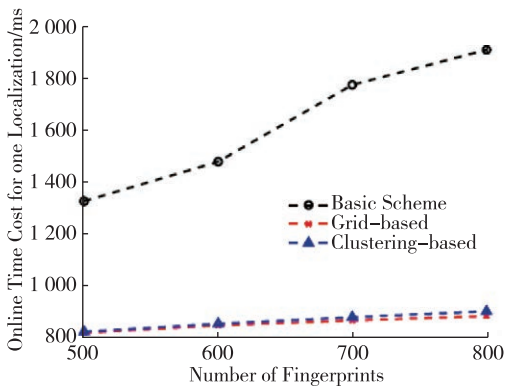

图 4 不同数据集大小下 3 种方案的时间消耗

Fig. 4 Online time costs of the schemes underdifferent sizes of fingerprint database

7.2 定位精度

除了效率以外还对定位精度的影响做了统计,显然决定能否准确定位的因素是能否找到距离用户测量的指纹最接近的数据库指纹.对每个实验都找了一个数据库的指纹当作用户测量的指纹,观察能否准确找到与他最接近的指纹.一开始,在网格的改进方案和聚类的改进方案中都不允许不同的网格或者分组中有交集元素.图 5 给出了 3 种方案的准确率,基础方案准确率最高,而基于聚类的方案要优于

基于网格的方案,这与我们在基于聚类的方案中所预测的一致.

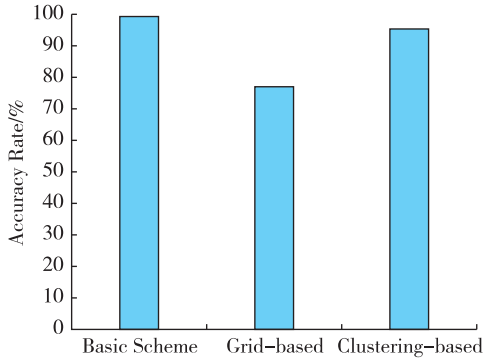

图5 不同方案寻找最接近指纹的准确率

Fig. 5 Accuracy rates of the proposed schemes forfinding the closest fingerprints

随后,允许网格或者分组中存在交集元素来改进准确率,我们尝试放宽目标指纹与中心指纹的距离阈值 θ 来决定一个指纹能否落入一个分组中.图6给出了随着 θ 的变化基于网格的方案和基于聚类的方案准确率的变化曲线.不过采用这种冗余设计会降低方案的时间效率,可以看到即使不采用这种设计,基于聚类的改进方案仍然有超过95%的准确率.

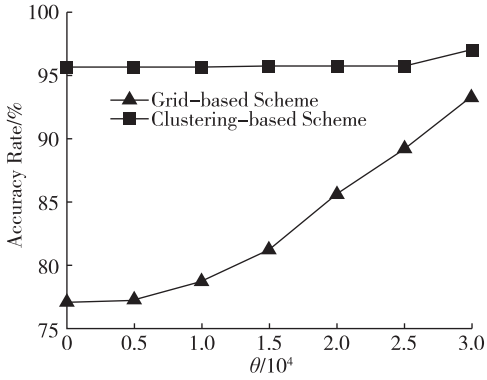

图6 允许指纹重复出现对寻找最接近指纹准确率的影响

Fig. 6 Effects of allowing overlaps among grids (clusters) on the accuracy rates for finding the closest fingerprints

7.3 对重复定位攻击的防御

第6部分提出了对于重复定位攻击的防御策略,我们在基于聚类的方案中实现了这一策略来评估对于定位精度的真实影响.首先模拟一个知道每个分组的非零特征与这个分组的指纹的攻击者,这个攻击者想要依靠这些信息来获取服务器的指纹.因为他知道每个组的指纹的非零特征,因此他只要知道这些特征的值就可以了.但是通过我们的防守策略,他可能会得到关于这些特征的错误值.

图7给出了使用防御策略后攻击者所需要发起的定位次数与共计有效的数量,发现当 $p<70\%$ 时80%以上的指纹需要攻击者进行100 000以上的定位才能破解,这对于攻击者来说并不现实.

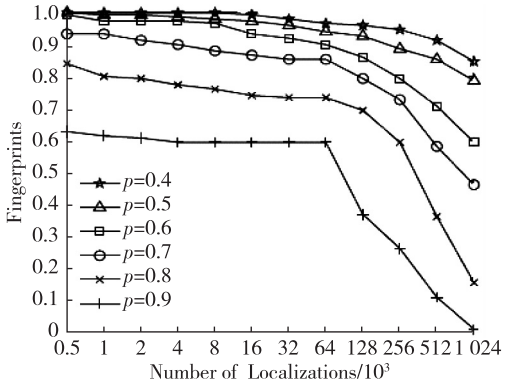

图7 破解一个指纹所需要的定位数量的累计分布函数

Fig. 7 CCDF of the number of localizations needed toreveal a fingerprint

8 总结

基于指纹的信号强度的定位是室内定位中最流行的技术,然而现有的工作并没有解决这个技术带来的隐私性问题:用户必须上传他的信息到服务器,服务器可以轻易定位追踪他.本文使用户可以在保护自己测量指纹数据的前提下借助服务器完成定位,而且在这个过程中服务商也不需要担心自己数据库的指纹泄露.首先提供了一个基础的基于ElGamal加密的方案来使用户和服务商在隐私安全的情况下进行指纹计算和匹配;随后为了减少时间开销和提高准确度,又提出了扩展的解决方案;最后为了加强服务器的安全性,给出了一个有效的针对重复定位攻击的防御方案.实验结果证明了本文方案在使用效率和定位精度上都具有很好的可用性.

参考文献

References

[1] Liu H B, Gan Y, Yang J, et al. Push the limit of WIFI based localization for smartphones[C]// Proceedings of the 18th Annual International Conference on Mobile Computingand Networking, 2012:305-316

[2] Azizyan M, Constandache I, Choudhury R R. Surround sense: Mobile phone localization via ambience fingerprinting [C] // Proceedings of the 15th Annual

International Conference on Mobile Computing and Networking, 2009: 261-272
[3] Tarzia S P, Dinda P A, Dick R P, et al. Indoor localizationwithout infrastructure using the acoustic backgroundspectrum[C]// Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services, 2011: 155-168
[4] ElGamal T. A public key cryptosystem and a signature scheme based on discrete logarithms [M]. Berlin: Springer, 1984: 469-472
[5] Wang T, Yang Y. Location privacy protection from RSS localizationsystem using antenna pattern synthesis[C]// Proceedings of IEEEINFOCOM, 2011: 2408-2416
[6] Gruteser M, Grunwald D. Anonymous usage of locationbased services through spatial and temporal cloaking[C]// International Conference on Mobile Systems, Applicationsand Services, 2003: 31-42
[7] Gedik B, Liu L. Location privacy in mobile systems: A personalized anonymization model [C] // IEEE International Conference on Distributed Computing Systems, 2005: 620-629
[8] Hoh B, Gruteser M, Xiong H, et al. Preservingprivacy in GPS traces via uncertainty-aware path cloaking[C]// ACM Conference on Computer and Communications Security, 2007: 161-171
[9] Kalnis P, Ghinita G, Mouratidis K, et al. Preventinglocation-based identity inference in anonymous spatial queries[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(12): 1719-1733
[10] Beresford A R, Stajano F. Location privacy in pervasive-computing[J]. IEEE Pervasive Computing, 2003, 2(1): 46-55
[11] Beresford A R, Stajano F. Mix zones: User privacy in locationawareservices[C]// Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops, 2004: 127-131
[12] Chow R, Golle P. Faking contextual data for fun, profit, andprivacy[C]// ACM Workshop on Privacy in the Electronic Society, 2009: 105-108
[13] Zhong S, Li L, Liu Y G, et al. Privacy-preserving location-based services for mobile users in wireless networks[R]. Yale Computer Science, Tech. Rep. YALEU/DCS/TR-1297, 2004
[14] Popa R A, Blumberg A J, Balakrishnan H, et al. Privacy and accountability for location-based aggregate statistics [C] // ACM Conference on Computer and Communications Security, 2011: 653-666
[15] Solanas A, Martinez-Balleste A. Privacy protection inlocation-based services through a public-key privacy homomorphism[C]// European Conference on Public Key Infrastructure: Theory and Practice, 2007: 362-368
[16] Swangmuang N, Krishnamurthy P V. On clustering RSS fingerprints for improving scalability of performance predictionof indoor positioning systems[C]// ACM International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments, 2008: 61-66
[17] Altintas B, Serif T. Improving RSS-based indoor positioningalgorithm via *k*-means clustering[C]// 11th European Wireless Conference 2011-Sustainable Wireless Technologies, 2011: 1-5
[18] Weiss J. Java cryptography extensions: Practical guide for programmers[M]. San Francisco, CA: Morgan Kaufmann Publishers Inc., 2004
[19] Yang Q, Pan S J, Zheng V W. Estimating location using wi-fi[J]. IEEE Intelligent Systems, 2008, 23(1): 8-13

Privacy-preserving in fingerprinting-based indoor localization

ZHANG Zhao[1] HUA Jingyu[1]
1 Department of Computer Science & Technology, Nanjing University, Nanjing 210023

Abstract Fingerprinting-based localization is one of the most popular indoor localization approaches. In the offline phase, the service provider measures the fingerprint, i.e., receives signal strength (RSS) samples from various access points (APs) at a number of knownlocations in the target space and stores them in a database. In the online phase, a user sends his location query with his current fingerprint measurement to the server, which will search for the closest fingerprintin the database. Although this approach has been studied for a long time, no existing work considers the privacy requirements for the two sides: the provider wants to protect thecollected fingerprints against the users; while the users want to protect their fingerprint measurements against the service provider to avoid location-leaking. In this paper, we aim to protect the privacy of the users and the service provider at the same time. We propose a privacy-preserving fingerprint matching scheme which uses a cryptographic technique to compute the distance

between the fingerprint measured by the user and the fingerprints in the database within the ciphertext space. We show that it well guarantees the privacy requirement of both the two sides in a single localization. To reduce its time overhead, we then present an improved scheme based on the grid division as well as three extensions at the cost of limited privacy loss. To enhance its security, we finally present an effective countermeasure against a special attack leveraging which malicious users could revealfingerprints on the server through repeated localizations. The extensive experiments with a public RSS-fingerprint dataset show that our proposal is fast enough for realtime localization and preserve the localization precision at the same time.

Key words fingerprinting-based localization; indoor localization; privacy preserving