轻量级位置感知推荐系统隐私保护框架

马鑫迪[1),2)] 李 辉[2)] 马建峰[1),2)] 习 宁[2)] 姜 奇[2)] 高 胜[3)] 卢 笛[1)]

[1)](西安电子科技大学计算机学院 西安 710071)

[2)](西安电子科技大学网络与信息安全学院 西安 710071)

[3)](中央财经大学信息学院 北京 102206)

摘 要 作为提供个性化位置服务的一种重要手段,高速、高效的位置感知推荐服务成为当前研究的热点.涉及多方参与的传统推荐流程存在着用户私密信息复制、盗取等安全威胁,给用户的隐私保护带来了新的挑战,尤其是当服务提供者将数据外包给第三方云平台时,隐私泄露问题会更加凸显.然而,现有的解决方案均存在推荐质量低、响应速度慢的问题.为解决上述问题,提出了一种轻量级的位置感知推荐系统隐私保护框架.利用该框架,服务提供者将随机处理后的历史评价信息外包给云平台,并通过安全协议在云平台的辅助下进行相似度信息的安全计算;同时,推荐用户利用可比较加密将其感兴趣的位置区域进行加密并发送给云平台进行请求服务,并通过安全协议实现推荐结果的安全预测.最后,通过在真实数据集中进行仿真调试,结果表明该框架能够在保证用户隐私安全的前提下,准确、高效地为用户推荐位置点.同时,与同态加密方案相比,该方案更加高效,能够更快速地响应用户的请求.

关键词 推荐系统;基于位置的服务;位置隐私;协同过滤;位置感知

中图法分类号 TP309 **DOI号** 10.11897/SP.J.1016.2017.01017

Lightweight Privacy-Preserving Framework for Location-Aware Recommender System

MA Xin-Di[1),2)] LI Hui[2)] MA Jian-Feng[1),2)] XI Ning[2)] JIANG Qi[2)] GAO Sheng[3)] LU Di[1)]

[1)](*School of Computer Science and Technology, Xidian University, Xi'an* 710071)

[2)](*School of Cyber Engineering, Xidian University, Xi'an* 710071)

[3)](*School of Information, Central University of Finance and Economics, Beijing* 102206)

Abstract As one of the most important method to provide individual location-based service for users, location-aware recommender system has recently experienced a rapid development and draw significant attention from the research community. Such problems are more prominent when service providers, who have limited computational and storage resources, leverage on cloud platforms to fit in with the tremendous number of service requirements and users. However, state-of-the-art work either suffers from an inaccurate recommendation quality or low efficiency. To address the problem, a lightweight privacy-preserving framework for location-aware recom-

收稿日期:2016-05-19;**在线出版日期**:2016-09-22.**本课题得到国家自然科学基金**(61202179,U1405255,61502368,61602537,61602357,61672413,U1509214,U1135002)、**国家"八六三"高技术研究发展计划项目基金**(2015AA016007)、**陕西省自然科学基金**(2015JQ6227,2016JM6005)、**国家111计划**(B16037)、**中央高校基本科研业务费**(JB150308,JB150309,JB161501)**资助**.**马鑫迪**,**男**,1989**年生**,**博士研究生**,**主要研究方向为隐私保护、网络安全**.E-mail:xdma1989@gmail.com.**李 辉**,**男**,1983**年生**,**博士**,**副教授**,**中国计算机学会**(CCF)**高级会员**,**主要研究方向为数据安全和隐私保护、数据库和数据挖掘理论和应用**.**马建峰**(**通信作者**),**男**,1963**年生**,**博士**,**教授**,**中国计算机学会**(CCF)**高级会员、会士**,**主要研究领域为密码学、计算机网络与信息安全**.E-mail:jfma@mail.xidian.edu.cn.**习 宁**,**男**,1986**年生**,**博士**,**讲师**,**中国计算机学会**(CCF)**会员**,**主要研究方向为服务计算、信息流控制**.**姜 奇**,**男**,1983**年生**,**博士**,**副教授**,**中国计算机学会**(CCF)**高级会员**,**主要研究方向为安全协议、无线网络安全**.**高 胜**,**男**,1987**年生**,**博士**,**讲师**,**中国计算机学会**(CCF)**会员**,**主要研究方向为金融信息安全、隐私计算**.**卢 笛**,**男**,1983**年生**,**博士**,**讲师**,**中国计算机学会**(CCF)**会员**,**主要研究方向为系统安全、云计算安全**.

mender system is proposed. Before requesting the service, service providers should firstly outsource the historical evaluation information, which is processed with random function, to cloud platform through the framework. Then, the similarities of information are computed securely with the help of cloud platform. When requesting the service, recommendation users encrypt the interesting area with comparable encryption and send the encrypted results to service providers. After that, the service providers will predict the recommendation results with the help of cloud platform through a carefully designed secure protocol. We have also theoretically proved that user information is private and will not be leaked during a recommendation. Finally, empirical results over a real-word dataset demonstrate that our framework can efficiently recommend POIs with a high degree of accuracy in a privacy-preserving manner. Compared with the existing work based on homomorphic encryption methods, our lightweight scheme is also more efficient and has a better user experience.

Keywords recommender system; location-based service; location privacy; collaborative filtering; location awareness

1 引 言

随着移动终端设备的不断普及和城市计算[1]的快速发展,基于位置的服务(Location-Based Service, LBS)已经得到广泛应用,如 Foursquare 推荐服务、Google 地图服务等.与此同时,由于有限的移动网络资源和硬件资源,服务类型与服务内容的日新月异为移动用户带来了严重的移动信息过载问题.如何从复杂的移动网络环境中发现用户真正感兴趣的信息资源,丰富并满足移动用户对服务的个性化需求,成为移动服务推荐领域亟待解决的技术难题.作为目前解决信息过滤和个性化服务问题的重要技术手段之一,推荐系统在 LBS 服务中发挥着越来越重要的作用.与传统的推荐系统相比,位置感知推荐系统除了考虑评价信息外,还需要考虑位置时空关联信息.

位置感知推荐系统中面临的主要挑战在于如何在大量的位置信息中为用户提供安全高效的推荐服务.但是,由于传统的位置感知推荐系统存在自身架构的弊端,并不能够在大规模数据集下为用户提供安全高效的推荐服务.首先,由于时空数据具有普遍存在和变化快速的特性,传统的位置感知推荐系统需要消耗大量的资源用于处理大规模时空数据的存储和计算问题,因此,许多 LBS 服务提供者开始向第三方云计算平台寻求帮助.通过将用户数据和推荐计算外包给云平台,LBS 服务提供者可以在保证服务质量的前提下,大大降低资源消耗,例如,Netflix 已经在 2015 年夏天关闭了其最后一个数据中心,并把所有的数据外包给亚马逊云计算平台存储和计算①.然而,云计算平台的引入又会带来一个新的挑战——隐私泄露.由于用户的请求信息和推荐结果中含有一定的隐私信息,如位置信息和偏好信息,如果不对其进行保护,云计算平台就可以获得这些信息,并跟踪用户或将用户的隐私信息发送给攻击者[2].因此,用户可能因担心其敏感信息泄露而拒绝使用此类服务,从而阻碍了推荐系统的发展,所以,如何保护用户的隐私信息在位置感知推荐系统发展过程中显得至关重要.

另外,位置感知推荐系统的隐私保护不仅仅包括用户的隐私,还包括服务提供者收集的用户历史评价信息和各个位置点的相似度信息的保护,这些数据被认为是服务提供者的私有"财产",为保证其利益,必须在加以保护后才可以外包给云计算平台.但是,截至目前为止,很少有研究工作在位置感知推荐系统中同时考虑这两个方面,如文献[3-5]仅考虑用户的隐私安全,并且其方案还存在服务质量差、推荐效率低的问题.

为解决上述问题,本文提出了一种轻量级的位置感知推荐系统隐私保护框架.在本框架中,服务提供者将数据和相关计算外包给云平台,从而降低其自身资源消耗,同时基于协同过滤算法,采用可交换加密和可比较加密实现云端数据安全存储和安全计

① Netflix shuts down its last data center, but it still runs a big it operation. http://arstechnica.com/information-technology/2015/08/Netflix-shuts-down-its-last-data-center-but-still-runs-a-big-it-operation/, Aug 2015

算，最终实现安全推荐. 最后，本文分析了该框架的安全性和效率，并进行了详实的理论证明和实际平台测试，结果表明：与采用同态加密相比，该轻量级框架占用的资源更少，能够更快速地响应用户的请求.

2　相关工作

近年来，位置感知推荐系统作为推荐服务中一大重要分支，越来越多的研究人员开始参与其中并展开相关工作[6]. 另外，随着用户对位置隐私关注度的增加，传统的未对用户隐私实施保护的位置感知推荐系统不再满足用户的需求，因此，学术界涌现出了大量的位置隐私保护方案和推荐系统隐私保护方案，这些方案主要通过匿名、差分隐私和密码学等技术手段实现.

2.1　位置隐私保护方案

针对位置隐私保护问题，研究人员已经提出大量的实施方案，但是这些方案只针对位置隐私，而不能直接应用于位置感知推荐系统中. 文献[7-9]提出多种基于匿名方案的位置隐私保护机制，这些机制主要通过混合区域、路径混淆等技术实现. 虽然此类匿名方案具有多样性，并且简单易用，但是该类方案均假设攻击者具有特定的背景知识，因此，在真实场景中并不能保护用户的位置隐私. 为弥补上述方案的缺点，文献[10-12]引入差分隐私保护机制用于保护用户的位置隐私，该类方案的优势在于不考虑攻击者的任何背景知识，但是其缺点在于在引入干扰噪声时，导致请求位置偏移，同时发起服务的请求范围扩大，容易给用户带来严重的通信开销及位置精度的损失，从而导致服务质量的降低. 另外，Shao 等人[13]提出了一种基于属性加密方案的细粒度位置隐私保护框架，该框架的优势在于不引入第三方的情况下，实现 LBS 访问策略及查询结果的机密性，但是该框架无法辅助服务提供者实现安全推荐计算. 作为一种通用的 SQL 加密查询框架，CryptDB[14]能够通过保序加密对加密的各位置点进行范围查询，但是它只支持查询，而不支持复杂计算，因此不能实现推荐结果的安全计算，所以，CryptDB 不能够应用在本文的场景中.

2.2　推荐系统隐私保护方案

截止目前，已经有大量的工作指出推荐系统为用户提供推荐服务的同时会造成隐私泄露的问题，第三方服务器或者攻击者在推荐过程中可以获得用户的偏好信息[15-16]. 同时，Staff[17]指出设计推荐系统的关键在于如何权衡隐私、效用及用户端的开销. 因此，服务推荐过程中的隐私保护成为当前的研究热点，如文献[3,18]中提出云计算环境下基于匿名方案的上下文感知推荐系统隐私保护机制；Guo 等人[19]提出了一种社交网络中基于信任的细粒度社交推荐隐私保护机制；Xin 等人[20]根据用户的隐私需求将其分为两类，对于少数“公开”用户，为保证其推荐的准确性，不采取措施对其隐私进行保护，而对于多数“私有”用户，则牺牲一定准确性并采取相应措施对其隐私进行保护；文献[21-23]则基于差分隐私提出了一种形式化的推荐系统隐私保护机制；文献[24-27]则提出基于密码学的方案保护用户在推荐系统中的隐私信息；Zhu 等人[28]基于移动终端 APP 流行度以及用户的隐私需求设计了个性化的具有隐私感知功能的智能 APP 推荐系统. 然而，上述工作均没有专门针对位置感知推荐系统设计相应的隐私保护机制，并且这些方案由于引入噪声或采用同态加密机制，均存在推荐质量低、响应速度慢的问题. 相比之下，本文设计的轻量级方案可以快速地响应用户的请求，并为其提供相对较高的服务质量.

3　预备知识

3.1　基本概念

3.1.1　协同过滤算法

协同过滤机制作为目前主流的推荐算法，被各大推荐服务提供商（如亚马逊）作为基本的推荐算法，因此，本文也选择协同过滤算法作为建立推荐系统隐私保护框架的基础. 基于物品的协同过滤算法作为较常用的协同过滤机制被应用在本文的方案中，首先，协同过滤算法假设一个用户集合 $U=\{u_1,\cdots,u_m\}$，$|U|=m$ 和一个位置点集合 $V=\{v_1,\cdots,v_n\}$，$|V|=n$，而且对于每一个位置点 v_i 都有一个三元组用于表示该位置点的属性信息：$A_{vi}=\{v_{iN},v_{ix},v_{iy}\}$，其中 v_{iN} 表示位置点 v_i 的 ID 标识，v_{ix}（或 v_{iy}）表示位置点 v_i 的纬度（或经度）. 每一个用户 u_j 可以向位置点集合中的各位置点发表评价信息，其中 r_{ij} 表示用户 u_j 对位置点 v_i 作出的评价信息. 当给定请求用户 u_q 时，协同过滤算法根据评价信息矩阵 $\boldsymbol{R}$ 及 u_q 的历史评价信息 $\boldsymbol{R}_u$ 为用户推荐出当前请求区域内其感兴趣的位置点.

为此，协同过滤算法首先计算各位置点之间的相似度信息矩阵 $\boldsymbol{Sim}$，其中 $Sim(v_p, v_q)$ 表示位置点 v_p 与 v_q 之间的相似度. 本文采用较流行的余弦相似度模型[29]计算相似度信息 $\boldsymbol{Sim}$：

$$Sim(v_p, v_q) = \frac{\boldsymbol{r}_{v_p} \times \boldsymbol{r}_{v_q}}{\|\boldsymbol{r}_{v_p}\| \times \|\boldsymbol{r}_{v_q}\|}.$$

然后，协同过滤算法按照如下公式预测位于用户请求范围内各个位置点的评价信息[30]：

$$P_{(u_q,i)} = \frac{\sum_{\ell \in \mathbf{V}_u} Sim(i,\ell) \times \boldsymbol{r}_\ell}{\sum_{\ell \in \mathbf{V}_u} |Sim(i,\ell)|}.$$

3.1.2 可交换加密

如果一种加密方案是可交换加密[31]，那么明文消息 m 被密钥 k_1 加密后再被密钥 k_2 加密，解密时用 k_1 对应的私钥对密文先解密，再用 k_2 对应的私钥进行解密，仍然可以得到正确的明文 m，即

$$D_{sk_2}(D_{sk_1}(E_{pk_2}(E_{pk_1}(m)))) = m.$$

本文引入可交换加密用于保护各位置点的属性信息（如 ID 标识信息和位置坐标信息），防止云计算平台和服务提供者在用户发送请求的过程中获得用户的推荐结果.

3.1.3 可比较加密

为解决保序加密（Order Preserving Encryption，OPE）在范围查询时与用户频繁交互的问题，Furukawa[32]提出了可比较加密的方案，采用可比较加密只需通过一轮交互就可以得到查询结果，同时还满足文献[32]中提出的弱安全性保证. 可比较加密方案通过 Der 和 Enc 加密函数对明文 num 加密分别生成令牌 $token$ 和密文 $ciph$. 因此，给定密文 $\{ciph, ciph'\}$，可比较加密可以通过如下方式比较其大小：

$$token = \mathrm{Der}(param, mkey, num),$$
$$ciph = \mathrm{Enc}(param, mkey, num),$$
$$ciph' = \mathrm{Enc}(param, mkey, num'),$$
$$\mathrm{Cmp}(param, ciph, ciph', token) = \begin{cases} 0, & num = num' \\ 1, & num > num' \\ 2, & num < num' \end{cases}.$$

3.2 系统架构及攻击模型

在描述位置感知推荐系统隐私保护框架前，本文首先对位置感知推荐系统定义如下.

定义 1. 给定用户 u_q 的位置坐标 (x_u, y_u) 及请求范围区间 $(\Delta x, \Delta y)$，位置感知推荐系统将基于 u_q 的历史评价信息 $\boldsymbol{R}_u$ 返回 u_q 对请求范围 $(x_u \pm \Delta x, y_u \pm \Delta y)$ 内各位置点的预测评价信息 $\boldsymbol{R}_p$.

3.2.1 系统架构

为保证云计算环境下位置感知推荐系统的隐私得到保护，所有的敏感信息都需要在密态数据下进行存储和计算. 因此，本文设计基于隐私保护的推荐系统架构如图 1 所示：

(1) **可信中心**（Trusted Authority，TA）. TA 独立于任何其他实体并被其他实体充分信任，主要负责为架构中的其他实体分发和管理密钥.

(2) **云平台**（Cloud Platform，CP）. CP 是架构中主要的存储和计算中心，主要负责管理、存储和计算数据.

(3) **服务提供者**（Service Provider，SP）. SP 拥有各位置点的属性信息以及定期收集到的用户历史评价信息. 然而，由于 SP 拥有有限的存储资源和计算资源，因此，必须将收集到的数据外包给 CP 存储和计算. 最后，SP 将会在 CP 的辅助下计算相似度信息和预测评价结果.

(4) **推荐用户**（Recommendation Users，RUs）. 在请求服务时，RUs 主要负责将其位置和请求区间发送给 CP 用于推荐计算.

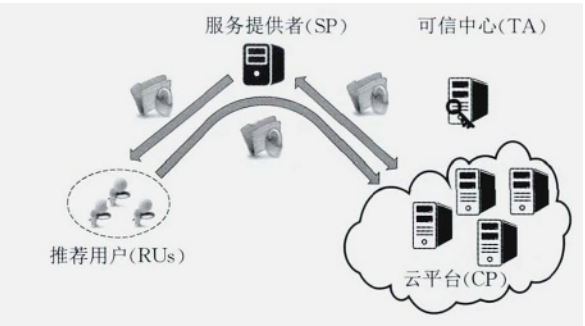


图 1 位置感知推荐系统架构

在上述系统架构下，各实体运行如下：

(1) SP 将定期收集到的历史评价信息 $\boldsymbol{R}_i$ 和位置属性信息 A_v 发送给 CP 存储，并在 CP 的辅助下，计算各位置点相似度信息矩阵 $\boldsymbol{Sim} \in \mathbf{R}^{n \times n}$.

(2) 当请求服务时，推荐用户 u_q 将其位置 (x_u, y_u) 及请求范围 $(\Delta x, \Delta y)$ 发送给 CP.

(3) CP 筛选出位于用户请求范围内的位置点 $\boldsymbol{V}'$，并对相关信息进行聚合计算，然后将聚合结果发给 SP.

(4) 基于 CP 的计算结果，SP 预测用户 u_q 对各位置点的评价信息 $\boldsymbol{R}_p$，并将 $\{(v'_N, v'_L), \boldsymbol{R}_p\}$ 发送给用户.

3.2.2 攻击模型

在位置感知推荐系统框架中，根据实际应用场景假设 SP 与 CP 是半可信的. 在计算各位置点间相

似度时，CP 作为主动攻击者期望获取 SP 采集到的历史评价信息 $\boldsymbol{R}$ 和相似度信息 $\boldsymbol{Sim}$，因此，这些数据作为 SP 的“私有财产”需要进行保护；同时，在 RUs 请求服务的过程中，CP 作为第三方云平台按照协议准确无误的执行用户与服务提供者之间的计算，但是 CP 也会作为主动攻击者期望获取用户的位置及其请求信息，另外，在此过程中，SP 也会按照协议正确执行计算，但是也会期望获得用户位置及请求信息. 最后，本文假设来自系统外部的攻击者作为被动攻击者也会发起攻击，通过监听信道等方式获取系统中所有的数据.

3.2.3　设计目标

本文致力于设计一种轻量级的位置感知推荐系统隐私保护框架，本框架为满足应用目标，需满足以下要求：

(1) 推荐质量. 本框架需实现有效推荐，使得推荐质量能够满足用户的需求.

(2) 推荐效率. 本框架需快速高效地为用户返回推荐结果，为用户提供更好的用户体验.

(3) 安全性. 本框架需满足隐私保护的需求，主要包括两个方面：首先，由于 SP 收集到的历史评价信息及相似度信息为私有数据，因此，在计算相似度时，CP 和外部攻击者不应获得 SP 的私有数据；其次，在 RUs 请求服务获得推荐结果的过程中，SP、CP 及外部攻击者不应获得用户的位置信息及推荐结果.

4　轻量级隐私保护框架

本节主要描述如何实现本文提出的轻量级位置感知推荐系统隐私保护框架，首先，本文将框架的实现过程按照协同过滤算法的执行流程分为两个阶段：基于隐私保护的相似度计算(阶段 1)与基于隐私保护的评价预测(阶段 2).

4.1　基于隐私保护的相似度计算(阶段 1)

在本阶段，SP 将收集到的历史评价信息及各位置点的属性信息外包给 CP，并在 CP 的辅助下计算相似度信息. 为保证 SP 私有数据在本阶段的安全性，本文分别采用可交换加密和可比较加密对位置点的属性信息和位置坐标信息进行加密，并设计一种轻量级的安全协议用于计算各位置点之间的相似度信息. 采用可交换加密的目的在于可以使用户最终安全地获得其请求范围内位置点的属性信息.

4.1.1　加密属性信息

为保证用户能够安全地获取位置点的属性信息，SP 首先将各位置点的属性信息进行扩展，将原三元组属性信息 $A_v=\{v_N,v_x,v_y\}$ 扩展为 $A'_v=\{(v_N,v_x,v_y),(v_x,v_y)\}=\{(v_N,v_L),v_L\}$，然后 SP 利用其可交换加密公钥 pk_s 对 (v_N,v_L) 进行加密，利用可比较加密对 v_L 进行加密，并将加密结果 $\{E_{pk_s}(v_N,v_L),\mathrm{Enc}(v_L)\}$ 发送给 CP.

4.1.2　轻量级相似度计算协议

在隐私保护框架中，SP 将历史评价信息和相似度信息作为私有数据，因此，在计算相似度信息时，CP 无法获得这些数据的明文信息. 详细的相似度计算协议如下：

步骤 1(@SP). 给定安全参数 k_1,k_2，大素数 p 和 α，并使 $|\alpha|=k_1$，选择大随机数 s 和 $n\times m$ 个随机数 c_{ij}，$i\in[1,n]$，$j\in[1,m]$，并使 $s\in Z_p$，$|c_{ij}|=k_2$. 对于任一 $r_{ij}\in\boldsymbol{R}_t$，执行如下计算[33]：

$$a_{ij}=s\times(r_{ij}\times\alpha+c_{ij})\bmod p.$$

本文用 $\boldsymbol{A}_t\in\mathbf{R}^{n\times m}$ 表示计算后得到的评价信息矩阵，$a_{ij}\in\boldsymbol{A}_t$，然后，SP 将 s^{-1}、p 和 α 保留，将 $\boldsymbol{A}_t$ 发送给 CP.

步骤 2(@CP). 当收到 $\boldsymbol{A}_t\in\mathbf{R}^{n\times m}$ 时，CP 首先将其整合到已有的历史评价信息矩阵 $\boldsymbol{A}\in\mathbf{R}^{n\times v'}$ 中，并对整合更新后的信息矩阵 $\boldsymbol{A}\in\mathbf{R}^{n\times v}$ $(v=m+v')$ 进行如下计算，$i\in[1,v]$，$j\in[1,n]$，$k\in[1,n]$ $(v\gg m\gg n)$：

$$\boldsymbol{B}=\boldsymbol{A}\times\boldsymbol{A}^{\mathrm{T}},$$

$$\begin{aligned}b_{jk}&=\sum_{i=1}^{v}a_{ji}\times a_{ik}\\&=\sum_{i=1}^{v}s^2(r_{ji}\times\alpha+c_{ji})(r_{ik}\times\alpha+c_{ik})\bmod p\\&=\sum_{i=1}^{v}s^2(r_{ji}r_{ik}\alpha^2+r_{ji}c_{ik}\alpha+c_{ji}r_{ik}\alpha+c_{ji}c_{ik})\bmod p.\end{aligned}$$

聚合过程如图 2 所示. 本文将上述计算结果表示为矩阵 $\boldsymbol{B}\in\mathbf{R}^{n\times n}$，$b_{jk}\in\boldsymbol{B}$. 经过计算后，原信息矩

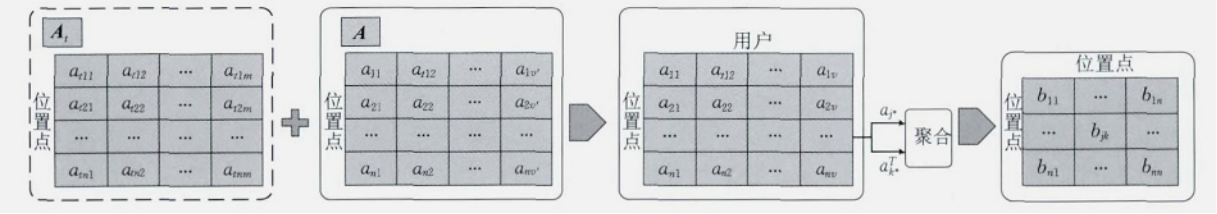


图 2　各位置点评价信息聚合(@CP)

阵维度得到压缩，当 CP 将矩阵 $\boldsymbol{B}$ 发送给 SP 时，SP 有足够的空间存储数据.

步骤 3(@SP). 当收到矩阵 $\boldsymbol{B}$ 时，SP 通过如下计算得到各位置点间的相似度信息，$j\in[1,n]$，$k\in[1,n]$：

$$d_{jk}=\frac{s^{-2}\times b_{jk}-s^{-2}\times b_{jk} \bmod \alpha^2}{\alpha^2}$$

$$=\sum_{i=1}^{v} r_{ij} r_{ik}.$$

因此，相似度信息可以通过如下公式计算：

$$sim_{jk}=\frac{d_{jk}}{\sqrt{d_{jj}}\times\sqrt{d_{kk}}}=\frac{\sum_{i=1}^{v} r_{ij} r_{ik}}{\sqrt{\sum_{i=1}^{v} r_{ij}^2}\times\sqrt{\sum_{i=1}^{v} r_{ik}^2}}.$$

为保证隐私保护框架能够快速地响应用户的请求，本文设计 SP 将相似度信息也外包给 CP，因此，SP 需要首先对相似度信息 $\boldsymbol{Sim}$ 进行如下处理：

(1) 选择 $n\times n$ 个随机数 w_{jk}，并使 $|w_{jk}|=k_2$，对于任一 sim_{jk} 进行如下计算：

$$f_{jk}=s\times(sim_{jk}\times\alpha+w_{jk}) \bmod p.$$

本文将处理后的相似度信息矩阵表示为 $\boldsymbol{F}\in\mathbf{R}^{n\times n}$，然后，SP 将矩阵 $\boldsymbol{F}$ 发送给 CP. 本协议的详细描述如图 3 所示，同时为保证结果的正确性，本文仍需要定义如下限制条件，若不保证下述条件，则在计算矩阵 $\boldsymbol{B}$ 和矩阵 $\boldsymbol{D}$ 的过程中求余运算时损失商倍数，后续计算时则不能按照上述规则进行计算，从而不能得到正确的计算结果：

$$\begin{cases} s^2(r_{ji}r_{ik}\alpha^2+r_{ji}c_{ik}\alpha+c_{ji}r_{ik}\alpha+c_{ji}c_{ik})<p \\ r_{ij}c_{ik}\alpha+r_{ik}c_{ij}\alpha+c_{ij}c_{ik}<\alpha^2 \end{cases}.$$

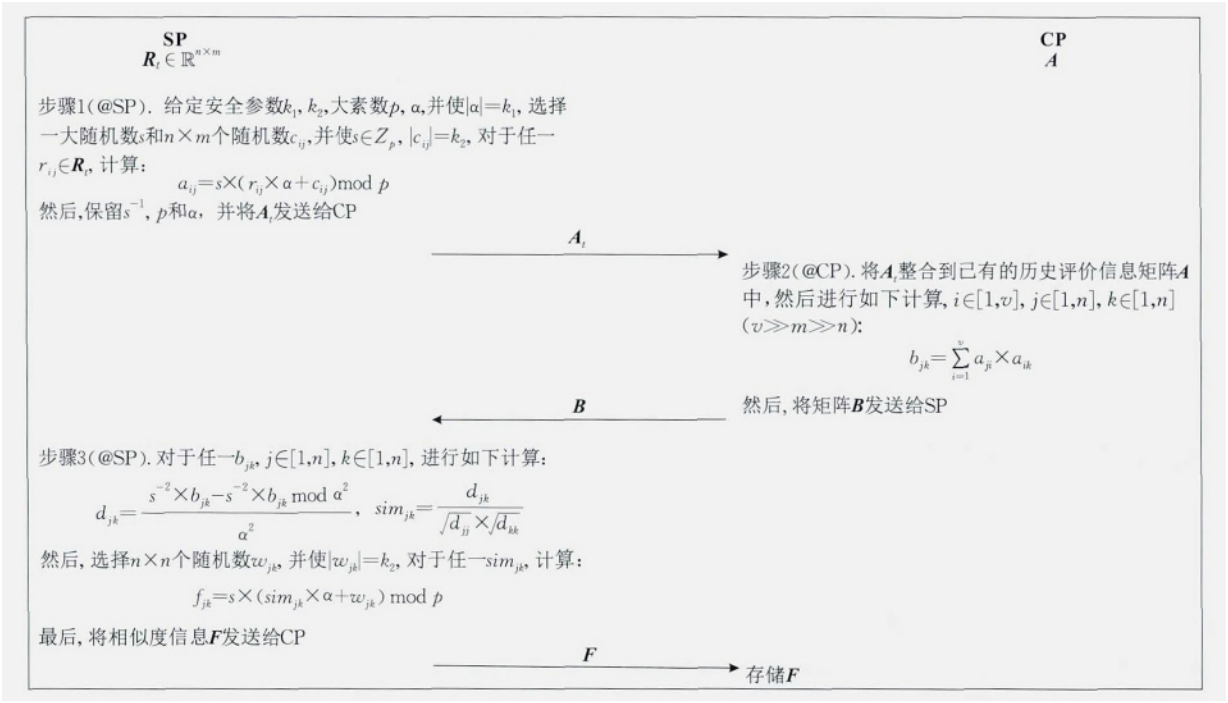


图 3 轻量级相似度计算协议

4.2 基于隐私保护的评价预测(阶段 2)

当发起服务请求时，用户需要将请求区域发送给 CP，同时为保证在推荐过程中，不泄露其对请求范围内位置点的预测评价信息，还需生成安全参数并一起发送给 CP. 详细评价预测协议如下：

步骤 1(@RUs). 向 CP 发起请求服务前，用户 u_q 按照如下步骤生成请求数据：

(1) u_q 对其请求区域 $\{x_u\pm\Delta x, y_u\pm\Delta y\}$ 采用可比较加密进行加密，得到 $\mathrm{Enc}(x_u\pm\Delta x, y_u\pm\Delta y)$ 和 $\mathrm{Der}(x_u\pm\Delta x, y_u\pm\Delta y)$.

(2) 当向 SP 注册为推荐用户时，u_q 将会收到 SP 向其发送的参数 α，同时给定安全参数 k_1、k_2、大素数 p，然后 u_q 选择大素数 β、γ，使得 $|\beta|=|\gamma|=k_1$，并生成大随机数 $s'\in Z_p$.

最后，u_q 将 $\{\mathrm{Enc}(x_u\pm\Delta x, y_u\pm\Delta y), \mathrm{Der}(x_u\pm\Delta x, y_u\pm\Delta y), s', \beta^2, \alpha\times\gamma, k_2\}$ 发送给 CP.

步骤 2(@CP). 当收到 $\{\mathrm{Enc}(x_u\pm\Delta x, y_u\pm\Delta y), \mathrm{Der}(x_u\pm\Delta x, y_u\pm\Delta y), s', \beta^2, \alpha\times\gamma, k_2\}$ 时，CP

首先用可比较加密筛选出位于用户请求范围内的位置点，然后，对这些位置点的相关信息进行聚合，并将聚合结果发送给 SP.

(1) 筛选位置点. 首先，CP 遍历所有的位置点坐标(v_{ix},v_{iy})，从中筛选出位于用户请求区域内的位置点，并将该位置点添加到集合 $\boldsymbol{H}$ 中. 筛选条件如下：

$$\begin{cases}\mathrm{Cmp}(param,\mathrm{Enc}(x_u-\Delta x),\mathrm{Enc}(v_{ix}),\\ \qquad \mathrm{Der}(x_u-\Delta x))=2\ \text{or}\ 0\\ \mathrm{Cmp}(param,\mathrm{Enc}(x_u+\Delta x),\mathrm{Enc}(v_{ix}),\\ \qquad \mathrm{Der}(x_u+\Delta x))=1\ \text{or}\ 0\\ \mathrm{Cmp}(param,\mathrm{Enc}(y_u-\Delta y),\mathrm{Enc}(v_{iy}),\\ \qquad \mathrm{Der}(y_u-\Delta y))=2\ \text{or}\ 0\\ \mathrm{Cmp}(param,\mathrm{Enc}(y_u+\Delta y),\mathrm{Enc}(v_{iy}),\\ \qquad \mathrm{Der}(y_u+\Delta y))=1\ \text{or}\ 0\end{cases}.$$

通过筛选之后，集合 $\boldsymbol{H}$ 中的所有位置点均位于用户的请求范围内，本文假设 $|\boldsymbol{H}|=h$. 然后，CP 用 u_q 的可交换加密公钥对集合 $\boldsymbol{H}$ 中的位置点属性信息再次加密，得到 $\boldsymbol{E}_{pk_u}(E_{pk_s}(v'_N,v'_L))$.

(2) 信息聚合. 首先，CP 从历史评价信息中提取用户 u_q 的历史评价信息 $\boldsymbol{A}_u$，$|\boldsymbol{A}_u|=\mathcal{L}$. $\boldsymbol{A}_u$ 中元素记为 $a_{u\ell}$，表示用户 u_q 对位置点 ℓ 的评价信息. 然后，CP 选择$\mathcal{L}$个随机数 $z_\ell,\ell\in[1,\mathcal{L}]$，并使 $|z_\ell|=k_2$，并对集合 $\boldsymbol{H}$ 中的任一元素执行如下计算：

$$\begin{aligned}q_{ui}&=\sum_{\ell=1}^{\mathcal{L}}s'(a_{u\ell}\times\beta^2+z_\ell\times\alpha\times\gamma)\times f_{i\ell}\\&=\sum_{\ell=1}^{\mathcal{L}}(s^2s'r_{u\ell}sim_{i\ell}\alpha^2\beta^2+s^2s'sim_{i\ell}c_{u\ell}\alpha\beta^2+\\&\quad s^2s'r_{u\ell}w_{i\ell}\alpha\beta^2+s^2s'c_{u\ell}w_{i\ell}\beta^2+\\&\quad ss'sim_{i\ell}z_\ell\alpha^2\gamma+ss'w_{i\ell}z_\ell\alpha^2\gamma)\bmod p,\\t_i&=\sum_{\ell=1}^{\mathcal{L}}f_{i\ell}=\sum_{\ell=1}^{\mathcal{L}}s(sim_{i\ell}\times\alpha+w_{i\ell})\bmod p.\end{aligned}$$

本文将聚合结果记为 $\boldsymbol{Q}\in\mathbf{R}^h$，$\boldsymbol{T}\in\mathbf{R}^h$，并将 $\{\boldsymbol{E}_{pk_u}(E_{pk_s}(v'_N,v'_L)),\boldsymbol{Q},\boldsymbol{T}\}$ 发送给 SP.

步骤 3(@SP). 当接收到 CP 发送的消息后，SP 进行如下计算：

(1) 首先，SP 利用其私钥 sk_s 对集合 $\boldsymbol{H}$ 中任一位置点属性信息进行内层解密，得到 $\boldsymbol{E}_{pk_u}(v'_N,v'_L)$.

(2) 其次，对于集合 $\boldsymbol{Q}$ 和 $\boldsymbol{T}$，SP 执行如下计算，$i\in[1,h]$：

$$r'_{ui}=\frac{s^{-2}q_{ui}-s^{-2}q_{ui}\bmod\alpha^2}{\alpha(s^{-1}t_i-s^{-1}t_i\bmod\alpha)}$$

$$=\frac{\sum_{\ell=1}^{\mathcal{L}}s'(r_{u\ell}sim_{i\ell}\beta^2+s^{-1}sim_{i\ell}z_\ell\gamma)\bmod p}{\sum_{\ell=1}^{\mathcal{L}}sim_{i\ell}}.$$

本文将上述计算结果记为集合 $\boldsymbol{R}'_{pu}$，由于 SP 无法获得 s'、β 等信息，所以其无法获得用户 u_q 的推荐结果. 最后，SP 将$\{\boldsymbol{E}_{pk_u}(v'_N,v'_L),\boldsymbol{R}'_{pu}\}$发送给 u_q.

步骤 4(@RUs). 当收到 SP 返回的数据后，用户 u_q 首先利用其私钥 sk_u 对集合 $\boldsymbol{H}$ 中任一位置点的属性信息进行解密，得到$\{v'_N,v'_L\}$. 为得到最终推荐结果，对集合 $\boldsymbol{H}$ 中任一位置点执行如下计算，$i\in[1,h]$：

$$r_{ui}=\frac{s'^{-1}r'_{ui}-s'^{-1}r'_{ui}\bmod\beta^2}{\beta^2}=\frac{\sum_{\ell=1}^{\mathcal{L}}r_{u\ell}sim_{i\ell}}{\sum_{\ell=1}^{\mathcal{L}}sim_{i\ell}}.$$

最后，用户 u_q 从筛选的位置点中选择预测评价最高的 k 个点作为其推荐结果. 本协议的详细描述如图 4 所示，同时为保证结果的正确性，本文仍需要定义如下限制条件，原因如阶段 1 中所述，如不保证下述条件，则在后续计算时不能按照上述规则进行，从而得不到正确结果：

$$\begin{cases}\sum_{\ell}^{L}s'(a_{u\ell}\times\beta^2+z_\ell\times\alpha\times\gamma)\times f_{i\ell}<p\\ \sum_{\ell}^{L}(s'sim_{i\ell}c_{u\ell}\alpha\beta^2+s'r_{u\ell}w_{i\ell}\alpha\beta^2+\\ \qquad s'c_{u\ell}w_{i\ell}\beta^2+s^{-1}s'w_{i\ell}z_\ell\alpha\gamma)<\alpha^2\\ \sum_{\ell}^{L}w_{il}<\alpha\\ \dfrac{\sum_{\ell=1}^{L}s^{-1}sim_{i\ell}z_\ell\gamma}{\sum_{\ell=1}^{L}sim_{i\ell}}<\beta^2\end{cases}.$$

5　框架安全与效率分析

本节主要从理论上分析该轻量级框架满足 3.2.3 节提出的安全性及效率要求. 首先，本文对该框架的安全性进行分析，然后，对系统提供服务时的存储和通信开销进行分析.

5.1　安全性分析

本节采用安全仿真模型[34-35]来证明本文框架的安全性，该模型常用于半可信环境下，证明多方协议的安全性. 安全仿真模型的含义在于：如果一条协议是安全的，那么协议中的任一方只会得到其对应

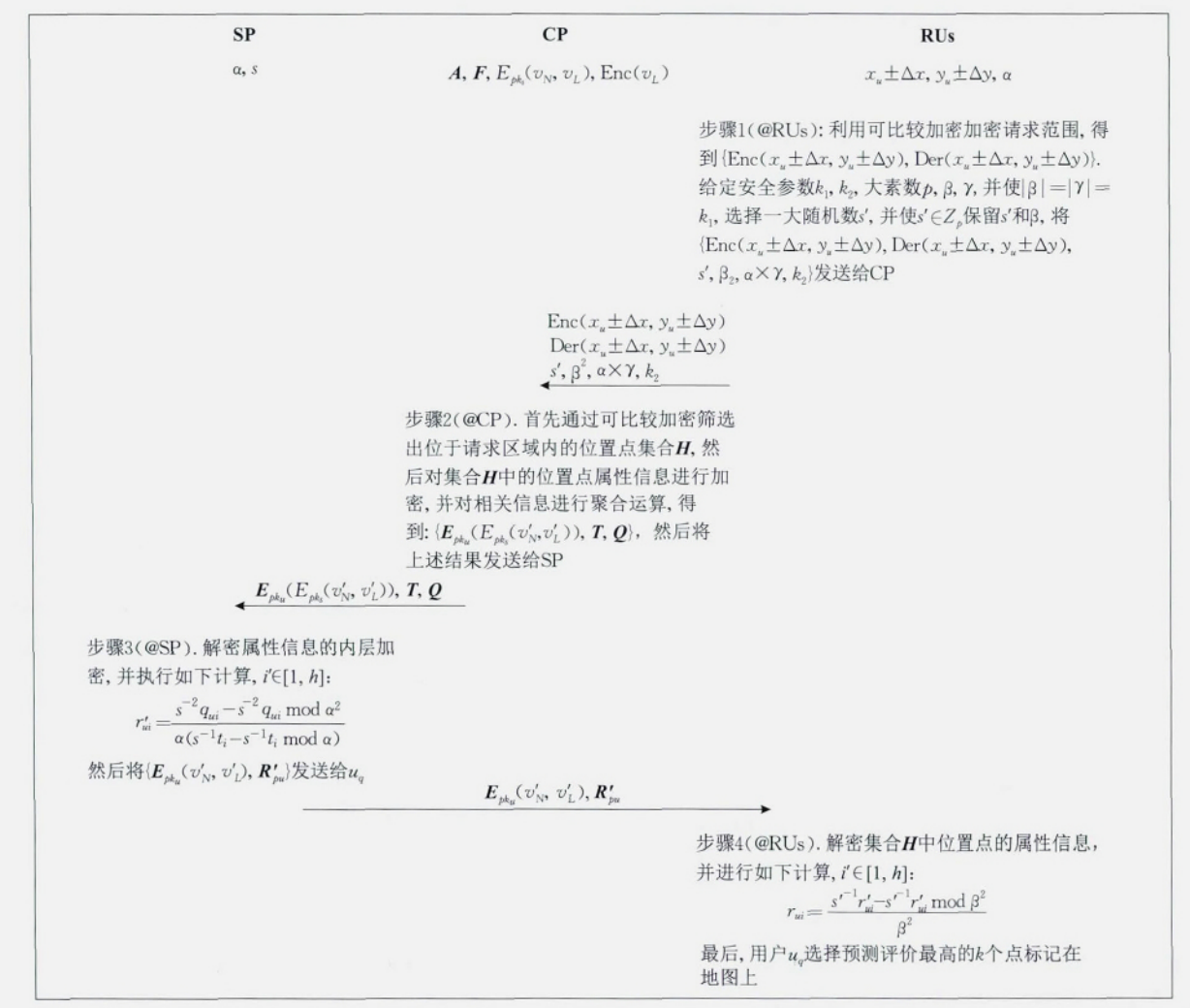


图 4 预测评价协议

的输入和输出内容,而无法得到执行计算过程中的其他任何内容.

定理 1. 在半可信模型下,本框架中阶段 1 和阶段 2 都是安全的.

5.2 存储与通信开销分析

本文假设在加密和随机化处理时,采用的安全参数最高为 1024 比特. 在阶段 1 中,SP 将历史评价信息和各位置点属性信息发送给 CP 存储,共需要 $O(m \times n)$ 通信开销,而 CP 将聚合运算结果发送给 SP 与 SP 将相似度信息发送给 CP 均需要 $O(n \times n)$ 通信开销,所以,在阶段 1 中,共需要 $O(n \times (m+2n))$ 通信开销. 另外,在阶段 1 中,SP 共消耗 $n \times (2n+m+2) \times 1024$ 比特的存储开销,与未采取加密及随机化处理方式相比,SP 需要消耗额外 $n \times (2n+m+2) \times 1023$ 比特的存储开销.

在阶段 2 中,用户请求服务时,只需要将感兴趣区域发送给 CP,消耗 $O(1)$ 通信开销,而在推荐的过程中,CP 将中间结果发给 SP 与 SP 将推荐结果发给用户均需要 $O(h)$ 通信开销,因此,共消耗 $O(2h)$ 通信开销. 另外,CP 需要消耗 $3h \times 1024$ 比特用于存储聚合计算结果,SP 需要消耗 $2h \times 1024$ 比特用于存储推荐结果,因此,在阶段 2 中,共需要消耗 $5h \times 1024$ 比特的存储开销,与未采用加密及随机化处理方式相比,需要消耗额外 $5h \times 1023$ 比特的存储开销.

6 性能评估

为验证本方案的高效性,本文设计实现该框架并在真实数据集上进行测试. 本文以 Archive 团队提取的 Foursquare 数据集(该数据集包含来自 2 153 471 个用户对 1 143 092 个位置点的 2 809 581

个评价信息①）为基础，随机选择来自 1400 个用户对 700 个位置点的 780 081 个评价信息进行测试. SP 与 CP 实验环境为 HP Pro 3380 MT，i5 CPU，8 GB 内存；RUs 的实验环境为华为荣耀 4A 手机终端，CPU 为 4 核 1.1 GHz，2 GB 内存.

6.1　推荐质量评估

为测试评估本框架的推荐质量，本文首先随机选取原始推荐数据的 50% 用于评估推荐结果的准确性，其余数据用于计算各位置点之间的相似度信息. 同时，本文随机选取一定位置区域作为用户感兴趣的区域，并通过测试程序发起服务请求.

在评估本框架推荐质量时，本文只需评估推荐结果中有多少位置点的的预测评价值与原始数据相同，因此，本文采用概率模型[36]度量原始数据与推荐结果之间误差发生的概率，公式如下：

$$P_d=\frac{\sum_{i=1}^{h}(|r_{pi}-r_i|=error)}{h},\ error=0,\cdots,5.$$

其中，h 表示位于用户请求区域内位置点的个数，r_{pi} 表示用户对位置点 v_i 的预测评价结果，r_i 表示原始数据中用户对位置点 v_i 的评价信息，$error$ 表示推荐结果误差. 测试结果如图 5 所示，从图中可以得到本文方案误差为 0 的概率高达 78.3%，误差为 4 和 5 的概率分别为 7.5%和 3.8%.

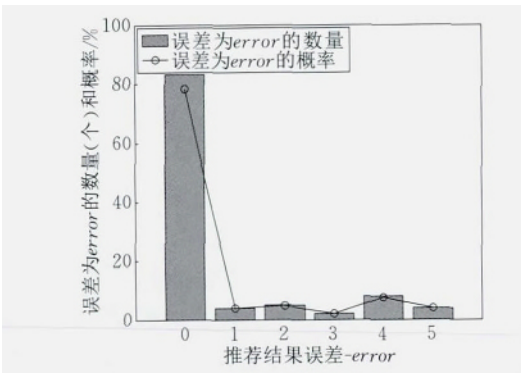


图 5　推荐质量评估

为证明本方案的有效性，本文将与文献[21,25]中的方案进行对比分析. 文献[21]中利用差分隐私方案对推荐系统中相关隐私信息进行保护，通过调节隐私开销 ε 实现不同程度的隐私保护，若调节 ε 能使其达到本文方案中的隐私保护效果，其推荐准确性将降低到 70%至 73%之间，然而当调节 ε 使其提高推荐准确性时，其隐私保护程度反而会相应降低. 文献[25]中利用加法同态加密对推荐系统中隐私信息进行保护，其可以达到与本文同等程度的隐私保护效果，但是其推荐准确性只有 65.4%. 因此，本文所述方案的推荐质量足够高，足以满足用户的服务请求.

6.2　推荐效率分析

在本框架中，主要有 3 个参量影响其效率性能：历史评价信息 $\boldsymbol{R}$ 中用户的数量 v，历史评价信息 $\boldsymbol{R}$ 中位置点的数量 n，用户 u_q 的请求范围区间（单位间隔为 1°）$(\Delta x,\Delta y)$. 因此，本文接下来主要评估上述参数对框架性能的影响.

在评估过程中，本文主要测试阶段 1 中预处理耗时和整个阶段 1 的耗时，测试阶段 2 中预处理耗时、整个阶段 2 的耗时及用户 u_q 最后计算推荐结果耗时.

本文首先测试历史评价信息 $\boldsymbol{R}$ 中用户的数量（v）对系统运行时间的影响，测试结果如图 6(a)、(b)所示. 从测试结果可以看出，随着用户数量的增加，阶段 1 的预处理时间及计算相似度信息的时间都会适当增加，而阶段 2 中的运行时间不受其影响. 这是因为当用户数量增加时，历史评价数据也会成倍增加，所以阶段 1 中预处理数据消耗时间会缓慢增加，而在阶段 1 中计算相似度信息需要遍历多次矩阵并进行大量的乘法操作，所以消耗时间增加得比较快；而阶段 2 中响应用户 u_q 的服务请求时，与历史评价信息中用户的数量无关，因此阶段 2 中运行时间及筛选出位置点的个数不变.

其次，本文测试历史评价信息 $\boldsymbol{R}$ 中位置点的数量（n）对系统运行时间的影响，测试结果如图 6(c)、(d)所示. 测试结果表明：随着位置点数量的增加，阶段 1 与阶段 2 中的时间消耗都会发生明显变化，原因在于当位置点数量增加时，历史评价信息会成倍增加，因此阶段 1 中时间消耗类似于图 6(a)中的变化，而在阶段 2 中，由于用户 u_q 预处理数据只是对其位置区域进行加密，并生成安全参数，因此，消耗时间与数据集 $\boldsymbol{R}$ 的大小无关，但是由于 CP 筛选位置点时遍历的数据会增多，所以阶段 2 中推荐消耗的时间会增加，而且随着筛选出位置点数目（h）的增加，整个推荐过程以及用户最后求解推荐结果的时间消耗都会明显增加，因此，阶段 2 的时间消耗会如图 6(d)中所示.

本文同时还测试了用户 u_q 请求范围 $(\Delta x,\Delta y)$ 的变化对系统运行时间的影响，测试结果如图 6(e)、(f)所示. 测试结果表明，阶段 1 的运行时间不受用户请求范围变化的影响，而阶段 2 的运行时间受请求范围变化的影响比较大，原因在于当用户请求范围变化时，阶段 1 中历史评价信息的数量不会发生变化，因此阶段 1 中的时间消耗不受其影响，而当用户的请求范围发生变化时，CP 在阶段 2 中筛选

① Umn/sarwat foursquare dataset. https://archive.org/details/201309_foursquare_dataset_umn, Oct 2013

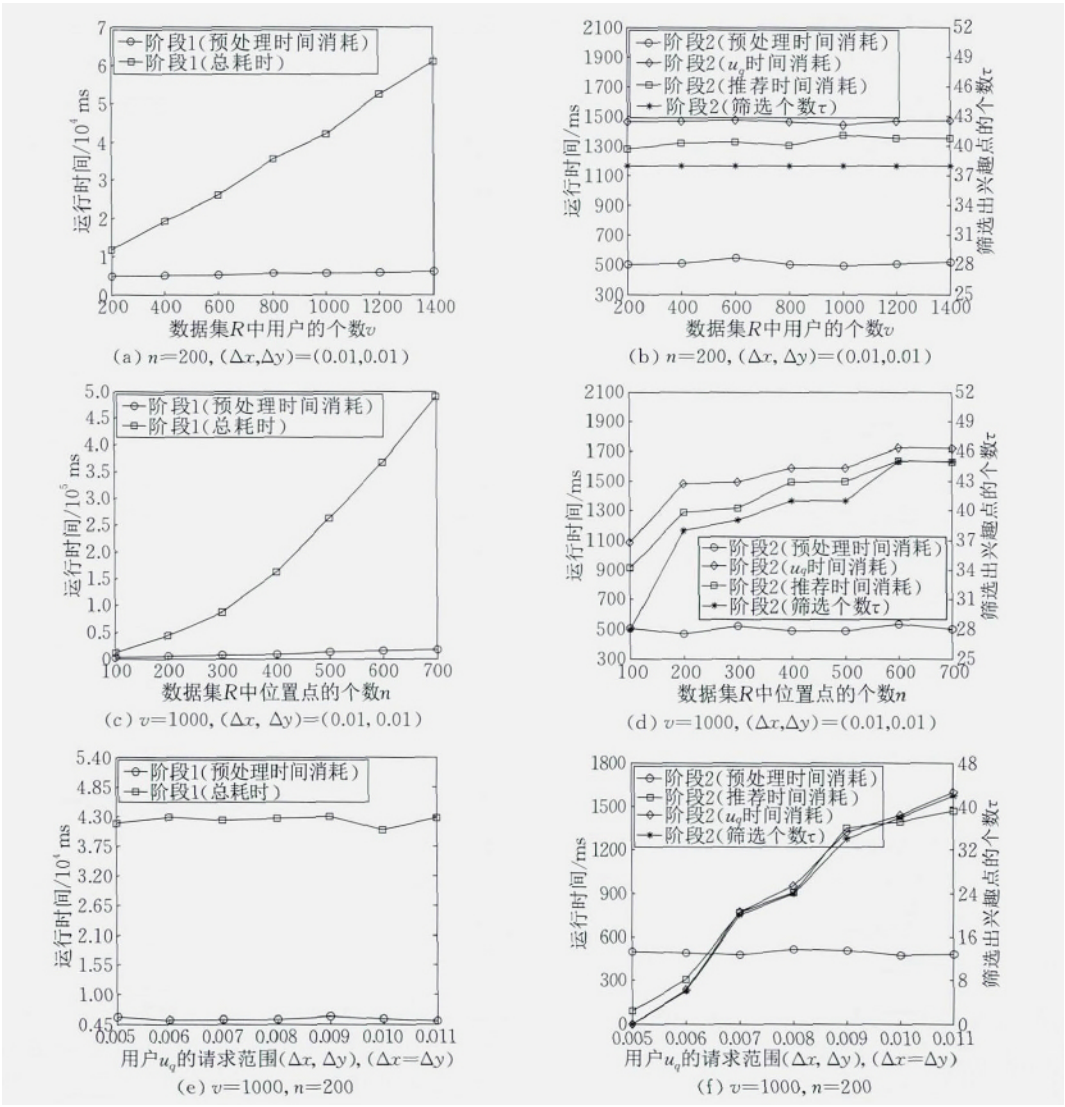


图 6 隐私保护框架测试评估

出的位置点数目(τ)会发生变化,如图 6(f)所示,随着用户请求范围的扩大,CP 筛选出的位置点数目也会增加,因此,阶段 2 中推荐过程消耗的时间和用户最后求解推荐结果消耗的时间也会增加.

6.3 方案对比分析

最后,本文与已有推荐系统隐私保护方案进行对比分析. 首先,本文通过实验与文献[37]中同态加密方案进行性能比较. 由于文献[37]中方案在相似度计算和用户请求推荐服务的过程中采用了同态加密方案,因此,文献[37]中方案会消耗更多的资源,系统耗时更长. 本文比较历史评价信息 R 中位置点的数量(n)对两种方案性能的影响,比较结果如图 7

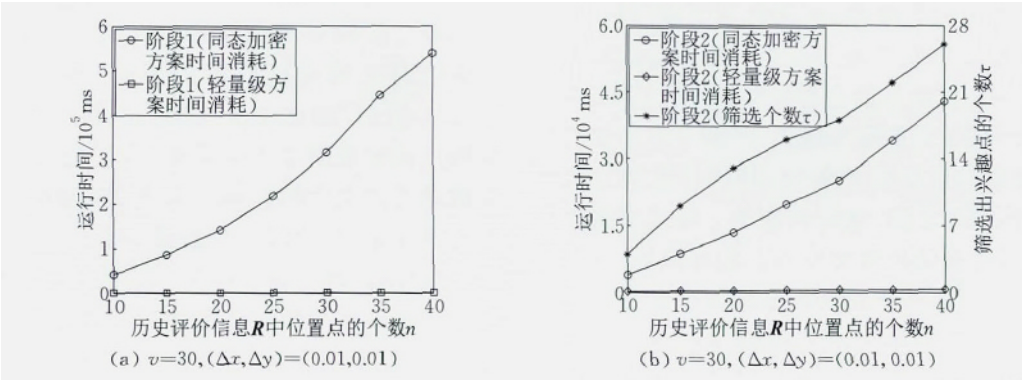


图 7 方案性能比较

所示. 测试结果表明：当历史评价信息 R 中位置点数量增加时，同态加密方案在阶段 1 和阶段 2 中消耗的时间都会明显增加，而本文的轻量级方案与同态加密方案相比，位置点个数在该数量级时，其消耗的时间可以忽略，因此，在性能方面，本文的方案优于文献[37]中的同态加密方案.

其次，本文与已有代表性方案在隐私保护效果、响应时间及推荐结果准确度等方面对比分析，分析结果如表 1 所述.

表 1　方案对比分析

方案	隐私保护机制	隐私保护强度	响应时间	推荐质量
SCIS'16[37]	同态加密	抵抗选择明文攻击	较慢(10 s 级)	高
ICWS'15[18]	模糊技术	不能抵抗背景知识攻击	快	低
TDSC'15[19]	假名技术	不能抵抗背景知识攻击	快(1 s 级)	高
PVLDB'15[22]	差分隐私	选择明文不可区分	较快	低
ICDM'14[23]	差分隐私	选择明文不可区分	慢(60 s 级)	低
TDP'15[25]	同态加密	抵抗选择明文攻击	较慢(10 s 级)	高
Globecom'14[26]	同态加密	抵抗选择明文攻击	慢(70 s 级)	高
本文方案	安全协议	抵抗选择明文攻击	快(1 s 级)	高

从上述对比结果可以看出，本文方案既在安全性方面达到了与同态加密同等级的隐私保护效果，又在效率方面达到了与假名技术同样的高效性，同时，还能保证高质量的推荐结果. 因此，综合隐私保护效果、推荐效率和推荐质量等方面考虑，本文方案与上述方案相比均具有一定的优势.

7　结　论

位置感知推荐系统中用户偏好信息的泄露严重威胁着用户的隐私安全，尤其是当服务提供者将数据外包给第三方云平台时，用户的隐私信息更容易被攻击者获取. 本文针对位置感知推荐系统中隐私泄露问题提出了一种轻量级隐私保护方案，采用安全协议的方式使得第三方云平台能够辅助服务提供者实现相似度信息的安全计算和推荐结果的安全预测. 另外，本文在真实数据的基础上，对该方案的性能进行了评估，并与采用同态加密的方案进行比较，结果表明本文提出的轻量级方案能够更快速地响应用户的请求.

参　考　文　献

[1] Zheng Y, Capra L, Wolfson O, et al. Urban computing: Concepts, methodologies, and applications. ACM Transactions on Intelligent Systems and Technology, 2014, 5(3): 38

[2] Levi A, Mokryn O, Diot C, et al. Finding a needle in a haystack of reviews: Cold start context-based hotel recommender system//Proceedings of the 6th ACM Conference on Recommender Systems. Dublin, Ireland, 2012: 115-122

[3] Celdran A H, Perez M G, Clemente F G, et al. PRECISE: Privacy-aware recommender based on context information for cloud service environments. IEEE Communications Magazine, 2014, 52(8): 90-96

[4] Huang J, Qi J Z, Xu Y B, et al. A privacy-enhancing model for location-based personalized recommendations. Distributed and Parallel Databases, 2015, 33(2): 253-276

[5] Scipioni M P. Towards privacy-aware location-based recommender systems//Proceedings of the 7th International Federation for Information Processing Summer School. Trento, Italy, 2011: 1-8

[6] Gao H J, Tang J L, Hu X, et al. Content-aware point of interest recommendation on location-based social networks//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, USA, 2015: 1721-1727

[7] Gao S, Ma J F, Shi W S, et al. TrPF: A trajectory privacy-preserving framework for participatory sensing. IEEE Transactions on Information Forensics and Security, 2013, 8(6): 874-887

[8] Niu B, Li Q H, Zhu X Y, et al. Enhancing privacy through caching in location-based services//Proceedings of the IEEE Conference on Computer Communications (INFOCOM). Hong Kong, China, 2015: 1017-1025

[9] Cicek A E, Nergiz M E, Saygin Y. Ensuring location diversity in privacy-preserving spatio-temporal data publishing. The Very Large Data Bases Journal, 2014, 23(4): 609-625

[10] Andres M E, Bordenabe N E, Chatzikokolakis K, et al. Geo-indistinguishability: Differential privacy for location-based systems//Proceedings of the 20th ACM SIGSAC Conference on Computer & Communications Security. Berlin, Germany, 2013: 901-914

[11] Xiao Y H, Xiong L. Protecting locations with differential privacy under temporal correlations//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. Denver, USA, 2015: 1298-1309

[12] To H, Ghinita G, Shahabi C. A framework for protecting worker location privacy in spatial crowd-sourcing. Proceedings of the Very Large Data Bases Endowment, 2014, 7(10): 919-930

[13] Shao J, Lu R X, Lin X D. FINE: A fine-grained privacy-preserving location-based service framework for mobile devices//Proceedings of the IEEE Conference on Computer Communications (INFOCOM). Toronto, Canada, 2014: 244-252

[14] Popa R A, Redfield C, Zeldovich N, et al. CryptDB: Processing queries on an encrypted database. Communications of the ACM, 2012, 55(9): 103-111

[15] Calandrino J A, Kilzer A, Narayanan A, et al. "You might also like:" Privacy risks of collaborative filtering//Proceedings of the IEEE Symposium on Security and Privacy (S&P). California, USA, 2011: 231-246

[16] Bhagat S, Weinsberg U, Ioannidis S, et al. Recommending with an agenda: Active learning of private attributes using matrix factorization//Proceedings of the 8th ACM Conference on Recommender Systems. Foster City, USA, 2014: 65-72

[17] Staff C. Recommendation algorithms, online privacy, and more. Communications of the ACM, 2009, 52(5): 10-11

[18] Zhu J, He P, Zheng Z, et al. A privacy-preserving QoS prediction framework for web service recommendation//Proceedings of the IEEE International Conference on Web Services. New York, USA, 2015: 241-248

[19] Guo L, Zhang C, Fang Y G. A trust-based privacy-preserving friend recommendation scheme for online social networks. IEEE Transactions on Dependable and Secure Computing, 2015, 12(4): 413-427

[20] Xin Y, Jaakkola T S. Controlling privacy in recommender systems//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2014: 2618-2626

[21] Jorgensen Z, Yu T. A privacy-preserving framework for personalized, social recommendations//Proceedings of the 17th International Conference on Extending Database Technology. Athens, Greece, 2014: 571-582

[22] Guerraoui R, Kermarrec A M, Patra R, et al. D2P: Distance-based differential privacy in recommenders. Proceedings of the Very Large Data Bases Endowment, 2015, 8(8): 862-873

[23] Shen Y, Jin H. Privacy-preserving personalized recommendation: An instance-based approach via differential privacy//Proceedings of the IEEE International Conference on Data Mining. Shenzhen, China, 2014: 540-549

[24] Liu B, Hengartner U. pTwitterRec: A rivacy-preserving personalized tweet recommendation framework//Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security. Kyoto, Japan, 2014: 365-376

[25] Samanthula B K, Cen L, Jiang W, et al. Privacy-preserving and efficient friend recommendation in online social networks. Transactions on Data Privacy, 2015, 8(2): 141-171

[26] Gong Y, Guo Y, Fang Y. A privacy-preserving task recommendation framework for mobile crowdsourcing//Proceedings of the IEEE Global Communications Conference. Austin, USA, 2014: 588-593

[27] Hoens T R, Blanton M, Steele A, et al. Reliable medical recommendation systems with patient privacy. ACM Transactions on Intelligent Systems and Technology, 2013, 4(4): 67

[28] Zhu H S, Xiong H, Ge Y, et al. Mobile app recommendations with security and privacy awareness//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA, 2014: 951-960

[29] Sarwat M, Leveandoski J J, Eldawy A, et al. LARS*: An efficient and scalable location-aware recommender system. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(6): 1384-1399

[30] Sarwar B M, Karypis G, Konstan J A, et al. Item-based collaborative filtering recommendation algorithms//Proceedings of the 10th International Conference on World Wide Web. Hong Kong, China, 2001: 285-295

[31] Weis S A. New Foundations for Efficient Authentication, Commutative Cryptography, and Private Disjointness Testing [Ph. D. dissertation]. Massachusetts Institute of Technology, Cambridge, USA, 2006

[32] Furukawa J. Request-based comparable encryption//Proceedings of the Computer Security-ESORICS. Egham, UK, 2013: 129-146

[33] Lu R X, Zhu H, Liu X M, et al. Toward efficient and privacy-preserving computing in big data era. IEEE Network, 2014, 28(4): 46-50

[34] Goldreich O. Foundations of Cryptography: Volume 2. Cambridge, UK: Cambridge University Press, 2009

[35] Bost R, Popa R A, Tu S, et al. Machine learning classification over encrypted data//Proceedings of the 22nd Annual Network and Distributed System Security Symposium. California, USA, 2015: 1-14

[36] Ye M, Yin P, Lee W C. Location recommendation for location-based social networks//Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. San Jose, USA, 2010: 458-461

[37] Ma X D, Li H, Ma J F, et al. APPLET: A privacy-preserving framework for location-aware recommender system. Science China Information Science, 2017, 60(9): 092101

附 录.

阶段 1 安全性证明.

在 3.2.3 节定义的安全性目标中,在计算各位置点相似度时,CP 作为半可信的攻击者期望获取 SP 的"私有数据",因此,在证明阶段 1 的安全性时,本文只需证明 CP 在执行

协议计算的过程中无法获得 SP 收集到的历史评价信息 $\boldsymbol{R}$ 即可.

首先,构造一个模拟器 S_{CP}^{A} 能够模拟一个视图与 CP 的真实视图 $V_{\mathrm{CP}}^{\pi A}(\boldsymbol{A};\boldsymbol{F})$ 是不可区分的,S_{CP}^{A} 执行如下操作:

(1) 生成一个 $n\times m$ 维的大随机数矩阵 $\overline{\boldsymbol{A}}$.

(2) 生成一个 $n\times m$ 维的矩阵 $\overline{\boldsymbol{F}}$.

(3) 按照协议执行计算并输出:$(\overline{\boldsymbol{A}},\overline{\boldsymbol{F}})$.

然后,定义如下等式:

① $H_0=V_{\mathrm{CP}}^{\pi A}(\boldsymbol{A})$

② $H_1=(\overline{\boldsymbol{A}};\boldsymbol{F})$

其中,H_0 表示输入为 $\boldsymbol{A}$ 时 CP 的视图,H_1 表示输入为 $\overline{\boldsymbol{A}}$,输出为 $\boldsymbol{F}$ 时的视图. 由于 SP 将历史评价信息 $\boldsymbol{R}$ 进行单向随机处理后生成矩阵 $\boldsymbol{A}$,因此,CP 无法区分 $\boldsymbol{A}$ 和 $\overline{\boldsymbol{A}}$,即 $(\boldsymbol{A})\overset{c}{\equiv}(\overline{\boldsymbol{A}})$,所以,$H_0\overset{c}{\equiv}H_1$. 综上所述,$V_{\mathrm{CP}}^{\pi A}\overset{c}{\equiv}S_{\mathrm{CP}}^{A}$.

另外,本框架分别采用可比较加密和可交换加密对各位置点属性和坐标进行加密,由两种加密方式的安全性可以得出攻击者不可能在概率多项式时间内解密数据. 因此,可以得出:CP 及其他攻击者在概率多项式时间内不能获得 SP 采集到的历史评价信息,阶段 1 在框架中是安全的.

阶段 2 安全性证明.

阶段 2 的安全性证明与阶段 1 类似,首先,构造一个模拟器 S_{uq} 能够模拟一个视图与 u_q 的真实视图 $V_{uq}^{\pi B}((x_u,y_u),(\Delta x,\Delta y),sk_u,param,mkey,\alpha,k_1,p;\beta,\gamma,s';\boldsymbol{E}_{pk_u}(v'_{\mathrm{N}},v'_{\mathrm{L}}),\boldsymbol{R}'_{pu})$ 是不可区分的,S_{uq} 执行如下操作:

(1) 生成随机参数 $\overline{\beta},\overline{\gamma}$ 和 $\overline{s'}$.

(2) 生成一组被可交换加密过的向量 $\overline{\boldsymbol{E}_{pk_u}}(v'_{\mathrm{N}},v'_{\mathrm{L}})$ 和一组随机向量 $\overline{\boldsymbol{R}'_{pu}}$.

(3) 按照协议执行计算并输出 $((x_u,y_u),(\Delta x,\Delta y),sk_u,param,mkey,\alpha,k_1,p;\overline{\beta},\overline{\gamma},\overline{s'};\overline{\boldsymbol{E}_{pk_u}}(v'_{\mathrm{N}},v'_{\mathrm{L}}),\overline{\boldsymbol{R}'_{pu}})$.

然后,定义如下等式:

① $H_0=V_{uq}^{\pi B}((x_u,y_u),(\Delta x,\Delta y),sk_u,param,mkey,\alpha,k_1,p)$.

② $H_1=((x_u,y_u),(\Delta x,\Delta y),sk_u,param,mkey,\alpha,k_1,p;\overline{\beta},\overline{\gamma},\overline{s'};\boldsymbol{E}_{pk_u}(v'_{\mathrm{N}},v'_{\mathrm{L}}),\boldsymbol{R}'_{pu})$.

③ $H_1=S_{uq}((x_u,y_u),(\Delta x,\Delta y),sk_u,param,mkey,\alpha,k_1,p;\boldsymbol{E}_{pk_u}(v'_{\mathrm{N}},v'_{\mathrm{L}}),\boldsymbol{R}'_{pu})$.

假设随机生成的 $(\overline{\beta},\overline{\gamma},\overline{s'})$ 与 (β,γ,s') 具有相同的概率分布,因此,在可比较加密的安全性保证下,H_0 和 H_1 是不可区分的,即 $H_0\overset{c}{\equiv}H_1$. 而且,若 $(\overline{\boldsymbol{E}_{pk_u}}(v'_{\mathrm{N}},v'_{\mathrm{L}}),\overline{\boldsymbol{R}'_{pu}})$ 与 $(\boldsymbol{E}_{pk_u}(v'_{\mathrm{N}},v'_{\mathrm{L}}),\boldsymbol{R}'_{pu})$ 具有相同的概率分布,由可交换加密的安全性可以得出 $H_1\overset{c}{\equiv}H_2$,因此,$V_{uq}^{\pi B}\overset{c}{\equiv}S_{uq}$.

其次,构造一个模拟器 S_{CP}^{B} 能够模拟一个视图与 CP 的真实视图 $V_{\mathrm{CP}}^{\pi B}(\boldsymbol{F},\boldsymbol{A},pk_u,E_{pk_s}(v_{\mathrm{N}},v_{\mathrm{L}}),\mathrm{Enc}(v_{\mathrm{L}}),param,k_2;\boldsymbol{Z};\mathrm{Enc}(x_u\pm\Delta x,y_u\pm\Delta y),\mathrm{Der}(x_u\pm\Delta x,y_u\pm\Delta y),s',\beta^2,\alpha\gamma)$ 是不可区分的,S_{CP}^{B} 执行如下操作:

(1) 生成一个随机区域并用可比较加密对其加密:$\overline{\mathrm{Enc}}(x_u\pm\Delta x,y_u\pm\Delta y),\overline{\mathrm{Der}}(x_u\pm\Delta x,y_u\pm\Delta y)$.

(2) 生成一个 $\mathcal{L}$ 长度的随机向量 $\overline{\boldsymbol{Z}}$ 和随机参数 $\overline{s'},\overline{\beta^2}$ 和 $\overline{\alpha\gamma}$.

(3) 以上述生成的随机参数和随机区域作为输入执行协议计算.

(4) 按照协议执行计算后输出 $(\boldsymbol{F},\boldsymbol{A},pk_u,E_{pk_s}(v_{\mathrm{N}},v_{\mathrm{L}}),\mathrm{Enc}(v_{\mathrm{L}}),param,k_2;\overline{\boldsymbol{Z}};\overline{\mathrm{Enc}}(x_u\pm\Delta x,y_u\pm\Delta y),\overline{\mathrm{Der}}(x_u\pm\Delta x,y_u\pm\Delta y),\overline{s'},\overline{\beta^2},\overline{\alpha\gamma})$.

然后,定义如下等式:

① $H_0=V_{\mathrm{CP}}^{\pi B}(\boldsymbol{F},\boldsymbol{A},pk_u,E_{pk_s}(v_{\mathrm{N}},v_{\mathrm{L}}),\mathrm{Enc}(v_{\mathrm{L}}),param,k_2)$.

② $H_1=(\boldsymbol{F},\boldsymbol{A},pk_u,E_{pk_s}(v_{\mathrm{N}},v_{\mathrm{L}}),\mathrm{Enc}(v_{\mathrm{L}}),param,k_2;\overline{\boldsymbol{Z}};\mathrm{Enc}(x_u\pm\Delta x,y_u\pm\Delta y),\mathrm{Der}(x_u\pm\Delta x,y_u\pm\Delta y),s',\beta^2,\alpha\gamma)$.

③ $H_2=(\boldsymbol{F},\boldsymbol{A},pk_u,E_{pk_s}(v_{\mathrm{N}},v_{\mathrm{L}}),\mathrm{Enc}(v_{\mathrm{L}}),param,k_2;\overline{\boldsymbol{Z}};\overline{\mathrm{Enc}}(x_u\pm\Delta x,y_u\pm\Delta y),\overline{\mathrm{Der}}(x_u\pm\Delta x,y_u\pm\Delta y),\overline{s'},\overline{\beta^2},\overline{\alpha\gamma})$.

若生成的随机向量 $\overline{\boldsymbol{Z}}$ 与 $\boldsymbol{Z}$ 具有相同的分布概率,则可以得到 $H_0\overset{c}{\equiv}H_1$. 同理,若 $(\overline{s'},\overline{\beta^2},\overline{\alpha\gamma})$ 与 $(s',\beta^2,\alpha\gamma)$ 具有相同的分布概率,同时在可比较加密的安全性保证下,可得 $H_1\overset{c}{\equiv}H_2$,因此,$V_{\mathrm{CP}}^{\pi B}\overset{c}{\equiv}S_{\mathrm{CP}}^{B}$.

最后,构造一个模拟器 S_{SP}^{B} 能够模拟一个视图与 SP 的真实视图 $V_{\mathrm{SP}}^{\pi B}(sk_s,s^{-1}\bmod p,\alpha;\boldsymbol{E}_{pk_u}(E_{pk_s}(v_{\mathrm{N}},v_{\mathrm{L}})),\boldsymbol{T},\boldsymbol{Q})$ 是不可区分的,S_{SP}^{B} 执行如下操作:

(1) 生成一个经可交换加密重加密后的随机向量:$\overline{\boldsymbol{E}_{pk_u}}(E_{pk_s}(v_{\mathrm{N}},v_{\mathrm{L}}))$.

(2) 生成随机向量 $\overline{\boldsymbol{Q}}$ 和 $\overline{\boldsymbol{T}}$.

(3) 以上述随机向量作为输入执行协议计算并输出:$(sk_s,s^{-1}\bmod p,\alpha;\overline{\boldsymbol{E}_{pk_u}}(E_{pk_s}(v_{\mathrm{N}},v_{\mathrm{L}})),\overline{\boldsymbol{T}},\overline{\boldsymbol{Q}})$.

然后,定义如下等式:

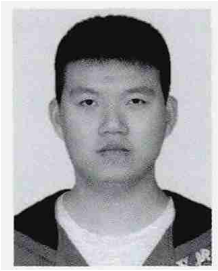
① $H_0=V_{\mathrm{SP}}^{\pi B}(sk_s,s^{-1}\bmod p,\alpha;\boldsymbol{E}_{pk_u}(E_{pk_s}(v_{\mathrm{N}},v_{\mathrm{L}})),\boldsymbol{T},\boldsymbol{Q})$.

② $H_1=(sk_s,s^{-1}\bmod p,\alpha;\overline{\boldsymbol{E}_{pk_u}}(E_{pk_s}(v_{\mathrm{N}},v_{\mathrm{L}})),\overline{\boldsymbol{T}},\overline{\boldsymbol{Q}})$.

假设随机生成的 $(\overline{\boldsymbol{T}},\overline{\boldsymbol{Q}})$ 与 $(\boldsymbol{T},\boldsymbol{Q})$ 具有相同的分布概率,同时在可交换加密的安全性保证下,本文可以保证 H_0 和 H_1 是不可区分的,即 $H_0\overset{c}{\equiv}H_1$,因此,$V_{\mathrm{SP}}^{\pi B}\overset{c}{\equiv}S_{\mathrm{SP}}^{B}$.

综上所述,没有恶意攻击者能够在概率多项式时间内区分模拟视图与真实视图,因此,在阶段 2 中,没有攻击者可以获得用户的隐私输入信息和推荐结果,同时,也没有攻击者可以在概率多项式时间内获得服务提供者的"私有数据".

证毕.



MA Xin-Di, born in 1989, Ph. D. candidate. His main research interests include privacy preserving and network security.

LI Hui, born in 1983, Ph. D., associate professor. His main research interests include security and privacy issues in data management, database and data mining principles and applications.

MA Jian-Feng, born in 1963, Ph. D., professor. His main research interests include cryptography, computer network and formation security.

XI Ning, born in 1986, Ph. D., lecturer. His main research interests include service computing and information-flow control.

JIANG Qi, born in 1983, Ph. D., associate professor. His main research interests include security protocols and wireless network security.

GAO Sheng, born in 1987, Ph. D., lecturer. His main research interests include finance information security and privacy computing.

LU Di, born in 1983, Ph. D., lecturer. His main research interests include system security and cloud computing security.

Background

With the development of urban computing and location-based services (LBSs), location-aware recommender systems have been widely used, providing us with an convenient way to experience life than ever before. Compared with traditional recommender systems, in addition to providing the recommendation ratings, location-aware recommender systems have to take into account both spatial-temporal and rating information.

The main challenge for location-aware recommender systems is how to securely and efficiently provide recommendations among a large number of point-of-interests (POIs). However, simply applying traditional recommendation techniques in an LBS may not be applicable for the following two reasons. First, as spatial data are ubiquitous and highly evolving, traditional recommender systems require a heavy toll for storing and computing user recommendations. Thus, many LBS providers have now turned to untrusted clouds for help. By moving user data and recommendation framework to the cloud, service providers can reduce the overhead of their computational resource while preserving the service quality. However, this technique has inevitably lead to another challenge, namely, privacy. Because user data and recommendation results include a certain amount of privacy information, such as locations and preferences, the cloud can easily infer what the users are interested in. Thus, the cloud may track users directly or release their preferential information to advertisers. As a result, users may be afraid that their sensitive information might be leaked to unauthorized attackers, which would be a huge barrier for the development and spread of recommender systems. Unfortunately, until now there have been limited research efforts or valuable contributions regarding this aspect.

In this paper, the authors design a lightweight framework to protect user privacy in location-aware recommender systems. Their proposed framework includes two lightweight protocols and employs the comparable and commutative encryption to protect the users' recommendation results during recommendations. Comprehensive analysis shows that the proposed framework is secure and can respond to user requests efficiently and effectively.

This research is supported by the National Natural Science Foundation of China (Grant Nos. 61202179, U1405255, 61502368, 61602537, 61602357, 61672413, U1509214, U1135002), the National High Technology Research and Development Program (863 Program) of China (Grant No. 2015AA016007), the Shaanxi Provincial Natural Science Foundation (Grant Nos. 2015JQ6227, 2016JM6005), the China 111 Project (Grant No. B16037), the Fundamental Research Funds for the Central Universities (Grant Nos. JB150308, JB150309, JB161501).