DOI:10.13530/j.cnki.jlis.182006

数字图书馆用户的行为偏好隐私保护框架*

吴宗大　谢　坚　郑城仁　周志峰　陈恩红

摘　要　针对新兴网络环境下数字图书馆用户的行为偏好隐私保护问题,设计实现了一个有效的方法框架。该方法框架的基本思想是:通过在可信客户端精心构造一系列“真假难辨”的伪行为,连同用户真行为一起,提交给不可信服务器端,“以假乱真”掩盖用户行为蕴含的敏感偏好。评估实验验证了该方法框架的有效性,即能在不损害数字图书馆服务的实用性、准确性和高效性的前提下,确保用户行为偏好隐私在不可信数字图书馆服务器端的安全性。该工作是针对数字图书馆用户行为偏好隐私保护问题的首次研究尝试,对搭建新网络环境下用户隐私安全的数字图书馆平台具有重要意义。图 5。表 1。参考文献 25。

关键词　数字图书馆　行为偏好　隐私保护

分类号　G250.76

A Framework for the Protection of User Behavior Preference Privacy of Digital Library

WU Zongda, XIE Jian, ZHENG Chengren, ZHOU Zhifeng & CHEN Enhong

ABSTRACT

Although providing great convenience for users, digital libraries result in users' serious concerns on personal privacy due to their more and more untrusted server-sides. In fact, users' privacy concerns have become one of the major obstacles to the development and application of digital libraries. In digital libraries, user privacy can be divided into data privacy and behavior privacy. Compared to data privacy, the protection of behavior privacy cannot be solved by using traditional privacy protection methods, because it is not allowed to change existing information services in digital libraries. Thus, it is more challenging to protect users' behavior privacy in digital libraries.

The purpose of this paper can be described as follows. Aiming at various kinds of online behaviors (i.e., service requests) issued by users in a digital library, we aim to construct a unified framework and model for behavior privacy protection, so as to break the limitations of traditional privacy protection methods when being applied to digital libraries, i.e., to ensure the security of various kinds of behavior privacy on the untrusted server-side, under the constraints of not changing the existing platform architecture and service algorithms of a digital library, and not compromising the accuracy and efficiency of information services supplied by the digital library.

* 本文系国家社会科学基金青年项目“数字图书馆用户的‘行为偏好隐私’保护方法研究”(编号:17CTQ011)的研究成果之一。(This paper is an outcome of the youth project “Research on the Protection of User Behavior Preference Privacy in a Digital Library” (No. 17CTQ011) supported by National Social Science Foundation of China.)

通信作者:吴宗大,Email:zongda1983@163.com, ORCID: 0000-0002-4292-3730 (Correspondence should be addressed to WU Zongda, Email:zongda1983@163.com, ORCID: 0000-0002-4292-3730)

In this paper, we first design a basic framework for user behavior privacy protection in a digital library. The basic idea of the framework is to lay a middleware (running at a trusted client, which is used to implement a privacy protection algorithm) between a library user interface (running at a trusted client) and the library services (running at the untrusted server); then, for a service request (i.e., a user behavior) issued by a user, the privacy algorithm would construct a group of high-quality dummy behaviors, and submit them together with the user behavior to the untrusted server-side, so as to cover up the sensitive preferences behind user's behaviors. Based on the framework, we then present a behavior privacy model, which formulates the constraints that ideal dummy behaviors should satisfy, to provide a reference for the privacy algorithm running at the client for the construction of dummy behaviors. Finally, we discuss the design and implementation of the privacy algorithm, under the model framework of users' behavior privacy protection.

Both theoretical analysis and experimental evaluation demonstrate the feasibility of the framework and model proposed in this paper, i.e., by constructing dummy behaviors of semantically irrelevant categories, the significance of users' sensitive preferences on the untrusted server-side can be reduced effectively (thereby, resulting in a good cover-up effect); and by constructing dummy behaviors of highly-similar feature distributions with user behaviors, it is difficult for attackers to rule out the dummy behaviors (thereby, resulting in a good mix-up effect).

This paper is the first research attempt to the protection of user behavior privacy in a digital library. The privacy framework proposed in this paper can ensure the security of user behaviors on the untrusted server-side, without compromising the availability, accuracy and efficiency of information services in a digital library, resulting in a positive significance to the development of a privacy-preserving digital library. However, this paper only describes a privacy framework at a high level of abstraction. In a digital library, there are various forms of behavior privacy (such as recommendation behavior and retrieval behavior). As the future work, we need to further study how to design and implement the corresponding privacy protection algorithm for each kind of user behavior. 5 figs. 1 tab. 25 refs.

KEY WORDS

Digital library. Behavior preference. Privacy protection.

0 引言

随着云计算等新兴网络信息技术的迅速发展,数字图书馆的应用领域得到不断延伸,已成为人们日常生活的重要组成部分。然而,在给用户带来巨大便利的同时,数字图书馆正变得越来越"不可信",从而引发数字图书馆用户对个人隐私安全的极度担忧[1-2]。用户隐私安全问题已成为制约数字图书馆发展与应用的主要障碍之一[3-4]。数字图书馆的用户隐私主要表现在以下两个方面[5]:①个人资料隐私,包括身份标识隐私(如身份证)和背景资料隐私(如职业);②行为偏好隐私,即使用图书馆服务时(如图书浏览服务、检索服务、推荐服务等),用户行为(服务请求)背后所蕴含的兴趣偏好隐私(如图书浏览行为蕴含用户偏好的图书类别)[6-7]。其中,资料隐私安全问题可通过数据加密技术较好地解决,即将用户资料加密后再存放到数字图书馆服务器中,这样即使它们不幸泄露,也难以被读懂[8-9]。然而,加密方法并不适用用户行为偏好隐私,因为图书馆服务需要服务器支

持,如果加密用户行为会使得服务器因无法“读懂”用户服务请求,而使得服务无法进行或有效完成[10-11]。为此,如何有效保护数字图书馆用户的行为偏好隐私安全,已成为一个至关重要的问题。

早期,图书馆领域的学者更多从法律角度研究图书馆用户隐私保护问题[4,12-14]。虽然制定隐私权相关的法律能在一定程度上保护用户隐私,但是并不能从根本上解决该问题,它更多地需要采用隐私保护技术来解决[5-6]。近年来,学者尝试从技术角度研究该问题[5-7],但已有方法还不够深入且缺乏系统,并且它们更多针对资料隐私,没有关注行为隐私。此外,针对不可信网络环境下的用户隐私安全问题,信息科学领域学者已给出了许多有效方法,代表性的有隐私加密技术、掩盖变换技术和匿名化技术。以下简要介绍这些方法的技术特点,并分析在数字图书馆中的应用局限性。①隐私加密技术是指通过加密变换,使得用户行为对服务器端不可见,以达到隐私保护的目的,代表性的有隐私信息检索技术[15-17]。该类技术不仅要求额外硬件和复杂算法的支持,且要求改变服务器端的服务算法,从而引起整个平台架构的改变,降低了方法在数字图书馆中的可用性。②敏感数据掩盖技术是指通过伪造数据或者使用一般化数据来掩盖涉及用户敏感偏好的行为数据[18-21]。由于改写了用户行为数据,该类方法对服务的准确性会造成一定负面影响,即其隐私保护需以牺牲服务质量为代价,难以满足数字图书馆的应用需求。③匿名化技术是用户隐私保护中广泛使用的一种技术,它通过隐藏或伪装用户身份标识,允许用户以不暴露身份的方式使用系统[22-24]。然而,匿名化隐私保护技术也受到了许多质疑。Josyula 等[25] 和 Narayanan 等[22] 分析了匿名化技术对隐私保护的不足,并给出实验证明。结果表明,通过匿名化技术收集的用户数据往往难以保证质量。更重要的是,数字图书馆一般要求用户必须实名登录后才能使用各项服务,所以,匿名化隐私保护技术难以有效地应用于数字图书馆。

综上所述,已有用户隐私保护技术并不是针对数字图书馆提出的,在实用性、准确性、安全性等方面仍无法满足数字图书馆的实际应用需求。理想的数字图书馆行为偏好隐私保护方法需要满足以下几个方面的要求:①确保用户行为隐私在不可信服务器端的安全性;②确保服务结果的准确性,即对比引入隐私保护方法的前后,用户获得的最终服务结果一致;③不损害数字图书馆信息服务的实用性,即隐私保护方法不改变服务器端的服务算法,不需要额外硬件支持,也不会对用户服务的使用效率产生显著影响。为此,本文的研究目标是:针对数字图书馆用户的各类行为,构建统一的行为偏好隐私保护框架模型,有效突破已有隐私保护技术在数字图书馆中的应用局限性,能在不改变现有数字图书馆平台架构、不改变现有图书服务算法、不改变图书服务准确性、基本不改变服务效率的前提下,确保各类用户行为偏好隐私在不可信服务器端的安全性。本文是针对数字图书馆用户行为偏好隐私保护的首次研究尝试,对构建新网络时代用户隐私安全的数字图书馆环境具有积极意义。

1 系统框架

数字图书馆提供的信息服务包括:图书浏览服务、检索服务、阅读服务、推荐服务等。在使用这些服务时,用户首先在客户端发起服务请求,服务器根据请求携带的数据,为用户提供相应图书服务。在数字图书馆中,服务器端是不可信的,它是攻击者的主要目标。因此,基于用户服务请求,不可信服务器可以分析出用户的兴趣偏好,从而导致用户行为隐私泄露。图 1 结合一个具体的图书浏览服务实例(即用户浏览“犯罪心理”相关图书),展示了本文采用的用户行为偏好隐私保护基本架构。该框架由一个

不可信服务器端和一组可信客户端组成,其一般化数据处理过程可简要描述如下。①客户端的"伪行为构造"部件通过分析用户行为 a_0,结合保存的"历史行为序列",综合考虑隐私安全性和服务高效性后,生成一组"真假难辨"的伪行为 $a_1,a_2,\cdots,a_n$,然后,将这些伪行为服务请求连同用户真行为服务请求 a_0,以随机次序提交给数字图书馆的服务器端,分别获取相应的图书服务。②客户端的"服务结果筛选"部件从服务器返回的结果集 $r_0,r_1,r_2,\cdots,r_n$(其中 r_k 是对应用户行为 a_k 的服务结果),筛选出对应用户真行为 a_0 的服务结果 r_0,同时丢弃其他伪服务结果 $r_1,r_2,\cdots,r_n$,然后,将 r_0 作为最终服务结果返回给客户端用户。

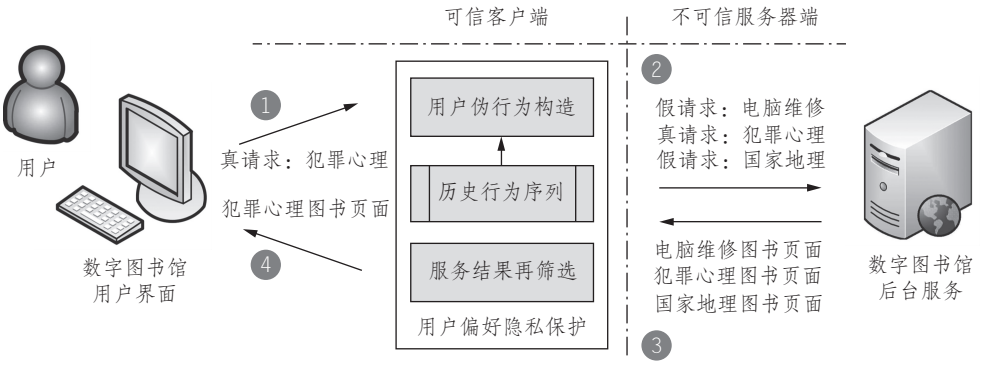

图 1　数字图书馆用户的行为偏好隐私保护架构

从上述系统架构可以看出:①用户隐私保护方法对服务器端后台服务和客户端外部用户均透明,因而,不改变现有数字图书馆平台架构;②服务器端所返回的结果必然是用户真实服务结果的超集,因而,不改变用户服务结果的准确性;③引入隐私保护方法所造成的服务性能下降程度线性相关于用户行为的数量,服务性能损失可控,不会显著降低服务效率。从上述系统架构还可以看出:"伪行为构造"部件伪造的服务请求至关重要。随机生成的伪行为通常无法实现"真假难辨"(容易被攻击者识别出来),难以有效保护用户行为隐私[16]。主要原因有以下几个方面。第一、用户当前行为本身会表现出富有规律的分布特征。例如,假定攻击者已获知用户身份为"一年级小学生"。给定两条图书浏览记录,如果它们对应图书的文体分别是"白话"和"文言",根据该语义特征,攻击者可以基本确定第二条浏览记录是伪造的。第二、用户当前行为与历史行为之间具有很强的分布特征连续性。例如,用户在一段时间内常常喜欢浏览或查询固定主题的图书(如数学),而随机生成的伪行为不具有这种特征连续性。攻击者据此可快速识别出伪行为。第三、用户各类别行为之间具有很强的语义关联性。例如,用户在一段时间内如果喜欢浏览"文学"类图书(浏览行为),则通常也喜欢阅读"文学"类图书(阅读行为)。而随机生成的各类伪行为之间不具有这种语义关联性。第四、为了保护用户隐私,伪行为不能与用户敏感图书类别(即用户不想被攻击者获知的兴趣图书类别)相关。例如,假定用户敏感类别为"犯罪心理",那么如果随机生成的伪行为也涉及该类别或与之高度相关的其他类别,是不合适的,因为攻击者不用识别出伪行为也能马上得出用户偏好"犯罪心理"的结论。

综上所述,理想伪行为应"真假难辨",能"以假乱真",掩盖用户敏感图书偏好,具体应满足以下两个条件:①与用户真实行为序列拥有高度相似的分布特征(具体包括单个行为之间的分布特征相似性、同类行为序列之间的连续

特征相似性,以及非同类行为序列之间的关联特征相似性),以有效地混淆用户提交的真实图书服务请求,使得攻击者无法排除这些伪服务请求(即实现"真假难辨"的目标);②与用户敏感偏好语义无关,以有效地掩盖用户敏感图书偏好,降低用户敏感偏好在不可信服务器端的暴露程度(即实现"以假乱真"的目标)。

2 隐私模型

基于前文系统框架,本小节定义一个面向数字图书馆用户的行为偏好隐私保护模型。据前文分析可知,理想的伪行为序列应"真假难辨"(即与用户行为序列特征相似),能"以假乱真",掩盖用户敏感图书偏好(即能有效降低敏感偏好在不可信服务器端的暴露程度)。为此,模型首先定义行为偏好(定义 3.1)和偏好频度(定义 3.2),据此进一步定义行为偏好暴露度(定义 3.3),以形式化度量伪行为对用户敏感偏好的掩盖效果(即"以假乱真"效果);然后,定义行为序列的分布特征、连续特征和关联特征(定义 3.4、3.6 和 3.8),以及各类特征之间的相似性(定义 3.5、3.7 和 3.9),以形式化度量伪行为序列对用户行为序列的混淆效果(即达到"真假难辨"效果);最后,基于前面两类度量,进一步形式化定义用户隐私保护方法所生成的理想的伪行为应满足的约束条件(定义 3.11)。

定义 3.1(行为偏好):$\mathbb{A}$表示所有可能行为组成的空间,$\mathbb{P}$表示所有可能偏好组成的空间。给定任意行为 $a\in\mathbb{A}$和任意偏好 $p\in\mathbb{P}$,它们之间的相关度可表示为函数 $Re(a,p):\mathbb{A}\times\mathbb{P}\rightarrow\mathbb{R}^+$($\mathbb{R}^+$表示正实数),则行为 a 背后所蕴含的偏好集合由所有与 a 相关的偏好组成,即:$P(a)=\{p\mid p\in\mathbb{P}\wedge Re(a,p)>\theta\}$。其中,阈值 θ 用于移除偏好空间$\mathbb{P}$中与行为 a 相关度较小的偏好(可简单设置为 0)。

定义 3.2(偏好频度):行为序列通常是指由用户在一段时间内发起的系列行为所组成的时间序列。任意偏好 $p\in\mathbb{P}$关于任意行为序列 A 的出现频度定义为:行为序列 A 中蕴含偏好 p 的行为数量,即:$Fr(p,A)=|\{a\mid a\in A\wedge p\in P(a)\}|$。显然,偏好频度值越大,偏好 p 与行为序列 A 的相关度越高。

定义 3.3(偏好暴露度):给定任意偏好 $p\in\mathbb{P}$ 和任意行为序列 A,偏好 p 关于行为序列 A 的暴露度可定义为

$$\exp(p,A)=Fr(p,A)/\sum_{a\in A}|P(a)|$$

给定任意行为序列集$\mathbb{A}'=\{A_1,A_2,\cdots,A_n\}$,偏好 p 关于行为序列集$\mathbb{A}'$的暴露度可定义为

$$\exp(p,\mathbb{A}')=\sum_{A\in\mathbb{A}'}Fr(p,A)/\sum_{A\in\mathbb{A}'}\sum_{a\in A}|P(a)|$$

显然,偏好 p 关于行为序列集$\mathbb{A}'$的暴露度越低,则攻击者从$\mathbb{A}'$分析出 p 的可能性就越低,即$\mathbb{A}'$对 p 的掩盖效果就越好。因此,可借助定义 3.3 形式化度量伪行为对用户敏感偏好的掩盖效果。据第 1 小节分析可知,用户行为序列特征包括:单个行为之间的分布特征、同类行为序列之间的连续特征,以及非同类行为序列之间的关联特征。为了实现"真假难辨"的目标,理想的伪行为要与用户行为拥有良好的整体特征相似度。为此,模型将分别定义行为序列的这些特征(分布特征、连续特征和关联特征)及其特征相似性,以有效度量伪行为序列关于用户真行为序列的整体特征相似度。以下,首先给出行为分布特征及其特征相似性定义。

定义 3.4(行为分布特征):给定任意行为 $a\in\mathbb{A}$,其某项分布特征仅与该行为自身相关,即行为分布特征函数可定义为 $F^p(a):\mathbb{A}\rightarrow\mathbb{R}^+$,它返回行为 a 的某项特征值。假定行为的可区分特征(即可区分不同行为的特征)共有 m 项,它们的函数分别记作:$F_1^p(a),F_2^p(a),\cdots,F_m^p(a)$,则行为 a 的分布特征可通过以下向量描述:$F^p(a)=[F_1^p(a),F_2^p(a),\cdots,F_m^p(a)]$。

定义 3.5(行为分布特征相似性):任意伪行为序列 $A_1=(a_1^1a_2^1\cdots a_n^1)$ 关于相应真实行为序列 $A_0=(a_1^0a_2^0\cdots a_n^0)$ 的分布特征相似性,可通过度量

它们相应分布特征向量之间的广义杰卡德相似性①来完成,即

$$\mathrm{sim}^p(A_1,A_0) = \frac{1}{n}\sum_{k\leqslant n} EJ(F^p(a_k^1),F^p(a_k^0)) = \frac{1}{n}\sum_{k\leqslant n}\frac{F^p(a_k^1)\cdot F^p(a_k^0)}{\|F^p(a_k^1)\|^2+\|F^p(a_k^0)\|^2-F^p(a_k^1)\cdot F^p(a_k^0)}$$

不同于分布特征,行为连续特征是指用户在一段时间内所发起的同类行为之间的相关性特征(如一段时间内用户可能会频繁地浏览某几本或某几类固定的图书),它主要用于捕获同一类别行为所表现出来特征的规律性。以下,形式化定义行为连续特征及其特征相似性。

定义 3.6(行为连续特征):给定任意行为 $a\in\mathbb{A}$和任意行为序列 $A\in 2^{\mathbb{A}}$,则行为 a 关于行为序列 A 的连续特征函数可定义为 $F^h(a,A):\mathbb{A}\times 2^{\mathbb{A}}\to\mathbb{R}^+$。记序列 A 的前 m 个行为所构成的子序列为 $A^m=(a_k)_{k=1}^m$,则行为序列 $A=(a_1a_2\cdots a_n)$ 连续特征可通过以下向量描述:$F^h(A)=[F^h(a_1,A^1),F^h(a_2,A^2),\cdots,F^h(a_n,A^n)]$。

定义 3.7(行为连续特征相似性):假定行为序列 A 的可区分连续特征(即可区分不同行为序列的连续特征)共有 m 项,它们的连续特征向量分别记作:$F_1^h(A),F_2^h(A),\cdots,F_m^h(A)$。任意伪行为序列 A_1 关于相应真实行为序列 A_0 的连续特征相似性(A_1 和 A_0 具有相同长度),可通过度量它们相应连续特征向量之间的广义杰卡德相似性来完成,即

$$\mathrm{sim}^h(A_1,A_0) = \frac{1}{m}\sum_{k\leqslant m} EJ(F_k^h(A_1),F_k^h(A_0))$$

数字图书馆为用户提供的服务可划分为以下几类:图书浏览服务、检索服务、阅读服务、推荐服务等。根据第 1 小节分析可知:除了同类用户行为序列会表现出富有规律的连续特征外,用户在一段时间内所发起的不同类别行为序列之间也会表现出富有规律的特征相关性(即行为关联特征)。以下,形式化定义行为关联特征及其特征相似性。

定义 3.8(行为关联特征):给定任意行为 $a\in\mathbb{A}_1$ 和其他类别的任意行为序列 $A\in 2^{\mathbb{A}_2}$($\mathbb{A}_1$ 与 $\mathbb{A}_2$ 属于不同类别,如浏览行为和阅读行为),则行为 a 关于行为序列 A 的关联特征函数可定义为 $F^r(a,A):\mathbb{A}_1\times 2^{\mathbb{A}_2}\to\mathbb{R}^+$。假定行为的可区分关联特征(即可区分不同行为的关联特征)共有 m 项,它们的函数分别记作:$F_1^r(a,A),F_2^r(a,A),\cdots,F_m^r(a,A)$,则行为 a 关于行为序列 A 的关联特征可通过以下向量描述:$F^r(a,A)=[F_1^r(a,A),F_2^r(a,A),\cdots,F_m^r(a,A)]$。

定义 3.9(行为关联特征相似性):对于任意伪行为序列 $A_1=(a_1^1a_2^1\cdots a_n^1)$,假定它关联的伪行为序列为 A_1^*。对于真行为序列 $A_0=(a_1^0a_2^0\cdots a_n^0)$,假定它关联的真行为序列为 A_0^*。伪行为序列 A_1 关于真行为序列 A_0 的关联特征相似性,可通过度量它们相应关联特征向量之间的广义杰卡德相似性来完成,即

$$\mathrm{sim}^r(A_1,A_0,A_1^*,A_0^*) = \frac{1}{n}\sum_{k\leqslant n} EJ(F^r(a_k^1,A_1^*),F^r(a_k^0,A_0^*))$$

综合定义 3.5(分布特征相似性)、定义 3.7(连续特征相似性)和定义 3.9(关联特征相似性)可进一步形式化度量伪行为序列对用户行为序列的混淆效果(即达到“真假难辨”效果)。

定义 3.10(行为特征相似性):任意给定用户图书行为序列 A_0,假定它由 n 个不同类别的行为子序列构成,即 $A_0=\langle A_1^0,A_2^0,\cdots,A_n^0\rangle$。给定匹配 A_0 的一个伪图书行为序列 A_1,即它同样由 n 个不同类别的行为子序列构成,即 $A_1=\langle A_1^1,A_2^1,\cdots,A_n^1\rangle$($A_k^1$ 对应 A_k^0),则 A_1 关于 A_0 的特征相似性可通过两者的分布特征相似性、连续特征相似性和关联特征相似性度量,即

$$\mathrm{sim}(A_0,A_1) = \sum_{k\leqslant n}\left(\frac{\mathrm{sim}^p(A_k^0,A_k^1)}{n}+\frac{\mathrm{sim}^h(A_k^0,A_k^1)}{n}+\sum_{h\neq k}\frac{\mathrm{sim}^r(A_k^0,A_k^1,A_h^0,A_h^1)}{n\cdot(n-1)}\right)$$

至此,基于定义 3.3(偏好暴露度度量)和定

① 见 https://en.wikipedia.org/wiki/Jaccard

义 3.10(行为特征相似性度量),以下进一步定义用户行为偏好隐私安全模型。

定义 3.11(用户行为隐私安全):任意给定一个用户行为序列A_0,假定它由 n 个不同类别的行为子序列构成。对于任意给定的一个伪行为序列集$\mathbb{A}'=\{A_1,A_2,\cdots,A_m\}$,其中,每个伪行为序列$A_k\in\mathbb{A}'$与$A_0$ 相匹配,即它同样由 n 个不同类别的伪行为子序列构成。如果$\mathbb{A}'$满足以下两个条件,则称:伪行为序列集$\mathbb{A}'$能有效地确保用户行为序列A_0 的(ω,ϖ)偏好隐私安全性。

(1)特征相似性。$\mathbb{A}'$中的各个伪行为序列 A_k 与用户真行为序列A_0 拥有高度相似特征(包括分布特征、连续特征和关联特征),即

$$\forall A_k\in\mathbb{A}'\rightarrow\mathrm{sim}(A_0,A_k)\geqslant\omega$$

其中,ω 为用户设定的阈值参数。该条件确保了伪行为序列与用户真行为序列之间的特征相似性,使得攻击者难以通过特征分析排除这些伪行为序列,从而使得用户行为序列被有效隐藏(即实现"真假难辨"目标)。

(2)偏好安全性。$\mathbb{A}'$中的伪行为序列能有效地降低用户敏感图书偏好集合$\mathbb{P}^*$(它由用户预先设定,且$\mathbb{P}^*\subset\mathbb{P}$)各个敏感偏好的暴露程度,即

$$\forall p^*\in\mathbb{P}^*\rightarrow\exp(p^*,A_0)/\exp(p^*,\{A_0\}\cup\mathbb{A}')\geqslant\varpi$$

其中,ϖ 为用户设定的阈值参数。该条件使得攻击者在没有排除伪行为序列的前提下,难以从用户提交的行为序列集中获知用户敏感图书偏好,从而确保用户行为偏好隐私的安全性(即实现"以假乱真"目标)。

以上 11 个定义构成了面向数字图书馆用户的行为偏好隐私模型。图 2 描述了这些定义之间的逻辑关系。在模型中,定义 3.1、3.4、3.6 和 3.8 尚缺少具体函数,而其他定义均直接或间接建立在这些函数之上。所以,如何准确地给出这些函数是模型求解的关键。

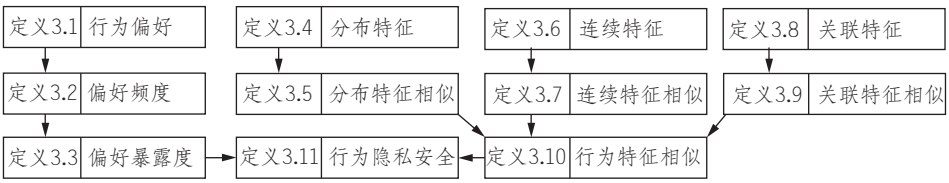

图 2 隐私模型定义间的逻辑关系

3 算法设计

基于第 2 小节给出的数字图书馆用户行为偏好隐私模型,本节讨论模型算法实现,以生成满足模型约束(定义 3.11)的伪图书服务请求序列。可以看出,第 2 小节的隐私模型是以图书行为序列为研究单位。然而,由于数字图书馆用户的行为序列是随时间动态增长的,算法难以一次性为用户行为序列构造生成完整的伪行为序列,为此,本节的实现算法将以单个行为为基本处理单位(即算法输入),即当用户在可信客户端发起一个当前图书服务请求时,隐私算法将结合客户端保存的历史图书行为序列(包括真行为序列和伪行为序列),构造生成一组相应的伪图书服务请求。然而,在处理用户的当前图书服务请求时,算法无法提前获知用户后续行为序列的内容,为此,算法采用"贪婪策略"(即在为用户当前图书服务请求构造伪请求时,只考虑当前构造的伪请求是否满足隐私模型的条件约束,而不考虑后续图书服务请求的伪造问题),但要求最终生成的伪行为序列能很好地满足定义 3.11 的条件约束。算法 1 描述了我们采用的算法基本实现方案。

首先,算法将为用户当前图书行为 a_t^0 获取一组图书分布特征相似但敏感偏好无关的伪图书行为候选者 A'(语句 6);然后,移除集合 A'中与用户行为 a_t^0 在图书连续特征上不够相似的伪

算法 1:为当前用户图书行为构造一组伪图书行为

输入:

①用户历史图书行为序列$A_0=\langle A_1^0,A_2^0,\cdots,A_n^0\rangle$

②历史伪图书行为序列 $A_1,A_2,\cdots,A_m$,其中,$A_k=\langle A_1^k,A_2^k,\cdots,A_n^k\rangle$

③用户当前图书行为 a_t^0 和用户敏感偏好集合$\mathbb{P}^*$

输出:一组伪图书行为 $a_t^1,a_t^2,\cdots,a_t^m$,分别关联伪行为序列 $A_1,A_2,\cdots,A_m$

01 设置$\mathbb{A}'$

02 FOREACH $A_k\in\{A_1,A_2,\cdots,A_m\}$ DO

03 基于用户设定的阈值参数,将变量 d_1、d_2 和 d_3 设置为合适的较小随机值

04 记用户行为序列 A_0 中与用户行为 a_t^0 拥有相同类别的行为子序列为 A_x^0

05 REPEAT

06 设置 $A'\leftarrow\{a\mid\forall p^*\in\mathbb{P}^*\rightarrow Re(a,p^*)\leqslant\theta\wedge EJ(F^p(a),F^p(a_t^0))\leqslant d_1\}$

/*获取一组与用户真行为拥有高度相似分布特征的伪行为候选者*/

07 FOREACH $a\in A'$DO

08 IF$\sum_q|F_q^h(a_t^0,A_x^0)-F_q^h(a,A_x^k)|>d_2$THEN 设置 $A'\leftarrow A'-\{a\}$

/*剔除候选集 A'中连续特征不相似的伪行为候选者*/

09 FOREACH $a\in A'$DO

10 IF$\sum_{i\neq x}EJ\ (F^r(a_t^0,A_i^0),F^r(a,A_i^k))>d_3$THEN 设置 $A'\leftarrow A'-\{a\}$

/*剔除候选集 A'中关联特征不相似的伪行为候选者*/

11 设置 $d_1\leftarrow d_1\cdot 2$,设置 $d_2\leftarrow d_2\cdot 2$,设置 $d_3\leftarrow d_3\cdot 2$

12 WHILE $A'\neq\varnothing$

13 从集合 A'中随机挑选 a_t^k,作为关联伪行为序列 A_x^k 的一个新的伪行为

14 设置$\mathbb{A}'\leftarrow\mathbb{A}'+\{a_t^k\}$

15 RETURN $\mathbb{A}'$

行为候选者(语句 7 至语句 8);接着,移除集合 A'中与用户行为 a_t^0 在图书关联特征上不够相似的伪行为候选者(语句 9 至语句 10);最后,从图书分布特征、连续特征和关联特征均与用户行为高度相似的伪行为候选者 A'中随机挑选一个行为 a_t^k,作为对应用户当前行为 a_t^0 的一个伪行为(语句 13)。在以上过程中,如果满足特征相似约束条件(即图书分布特征相似、连续特征相似,以及关联特征相似)的伪行为候选者 A'为空(语句 12),算法将松散特征相似约束条件(语句 11),然后,重新搜索一组满足条件的伪行为候选者。还可以看出,算法 1 的输出是不确定的,即对于同样的输入,两次运行会得到不同的输出,因为语句 3 和语句 13 进行了随机操作。这种做法是为了更好地保证算法的安全性(具体见第 5 小节的分析讨论)。为了方便算法描述,算法 1 还假定伪行为候选集有解(即语句 12 的约束条件不会永远为真)。从算法 1 可以看出,由于历史伪图书行为序列与用户真实图书行为序列通常具有相同长度,因此,算法 1 的时间复杂度为 $O(m\cdot|A_0|)$。

4 实验评估

4.1 实验准备

本小节旨在验证前述图书行为偏好隐私保护模型的有效性。数字图书馆为读者提供的信息服务形式多样。为了简化实验设计,这里只考虑相对简单的图书浏览服务和图书阅读服务。为了获得实验数据,我们收集了温州大学图书馆

100 名读者近年来的图书服务记录,并为每位读者精心挑选了 1 000 条浏览记录和1 000条阅读记录(即用户行为序列长度为 2 000,由相同长度的阅读行为子序列和浏览行为子序列构成)。

前文给出的仅是一个针对图书馆用户的行为偏好隐私保护框架模型,其实际运行还依赖于行为偏好相关度、行为分布特征、行为连续特征和行为关联特征等函数的准确实现。为此,需要研究在仅考虑图书浏览行为和阅读行为时,上述四类函数的具体实现方法。笔者注意到用户的一条图书浏览或阅读记录(即浏览行为或阅读行为)通常对应着一本具体图书,为此,借助于用户行为蕴含的具体图书信息,可构建上述四类函数。为了构建行为偏好相关度函数(定义 3.1),挑选"中图法图书分类目录"中处于次顶层的图书目录(如 B0 哲学理论、B1 世界哲学、D0 政治理论等)组建行为偏好空间$\mathbb{P}$,然后,以图书分类目录为中间媒介构建行为偏好相关度函数(行为对应图书,偏好对应图书目录)。对于行为分布特征函数(定义 3.4),主要考虑了图书长度、文体、价格、语言等基本特征。对于行为连续特征函数(定义 3.6)和关联特征函数(定义 3.8),主要考虑了行为频度和偏好频度两类特征。表 1 给出了这些函数的实现方法。

表 1 图书行为函数的具体实现方法

行为函数类别	行为函数	行为函数值含义
行为偏好相关度	$Re(a,p)$:相关度	0(行为 a 不属于偏好目录 p);1(否则)
行为分布特征	$F_1^p(a)$:图书长度	0(小于 0.5 万字);1(0.5 万至 1 万)…
	$F_2^p(a)$:图书文体	0(小说);1(散文);2(论文)…
	$F_3^p(a)$:图书语言	0(中文);1(英文);2(日文)…
	$F_4^p(a)$:图书价格	0(小于 10 元);1(10 元至 20 元)…
行为连续特征	$F_1^h(a,A)$:图书频度	序列 A 中出现图书(行为)a 的次数
	$F_2^h(a,A)$:偏好频度	序列 A 中属于偏好(目录)p 的行为数量
行为关联特征	$F_1^r(a,A)$:图书频度	序列 A 中出现图书(行为)a 的次数
	$F_2^r(a,A)$:偏好频度	序列 A 中属于偏好(目录)p 的行为数量

此外,实验还引入了随机方法以进行比较。随机方法中,伪阅读记录(或伪浏览记录)所关联的图书是随机选取的,但要求伪行为序列长度与用户真行为序列一致;各个伪行为与其对应的真行为类别一致。这里没有与引言中提到的其他隐私方法进行比较,这是因为这些方法并不是针对数字图书馆提出的,它们与本文方法建立在不同的系统模型和隐私模型上,难以与本文方法直接进行比较(作为替代,第 5 小节进行了定性分析比较)。实验中,所有算法都是用 Java 语言完成。实验是在配置为 Intel Core 2 Duo 3GHz CPU 和最大工作内存为 2GB 的 Java 虚拟机器(版本 1.7.0 07)上执行。

4.2 实验结果

实验一旨在评估本文方法所产生的伪行为对用户敏感偏好的掩盖效果。这里使用"偏好暴露度"(参考定义 3.3 构建),以度量敏感偏好$\mathbb{P}^*$关于行为序列集$\mathbb{A}'=\{A_0,A_1,\cdots,A_m\}$的暴露度,即$\max_{p^*\in\mathbb{P}^*}(\exp(p^*,A_0)/\exp(p^*,\mathbb{A}'))$。显然,度量值越小越好,因为它意味着攻击者越难从行为序列集$\mathbb{A}'$中直接猜测用户敏感图书偏好。该度量主要取决于敏感偏好数量和构造的伪行为序列数量。实验中,行为序列长度固定为 2 000。实验评估结果如图 3 所示,其中,子图左下角指示预先设定的用户敏感偏好数量(M=1、M=3 和 M=5)。从图 3 可以看出,本文方法

生成的伪行为序列能有效地改善敏感偏好的暴露程度,并且这种改善效果基本上与伪行为序列数量正相关,不会随着敏感偏好数量的改变而明显改变。相比于本文方法,随机方法生成的伪行为虽然也能在一定程度上降低敏感偏好的暴露程度,但稳定性较差(即不与伪行为序列数量正相关),并且其效果会随着敏感偏好数量的增加而变差。后续实验也表明:随机方法所生成的伪行为序列与用户真实行为序列的特征相似性很差,使得它们容易被攻击者排除,难以有效地保护用户敏感偏好。

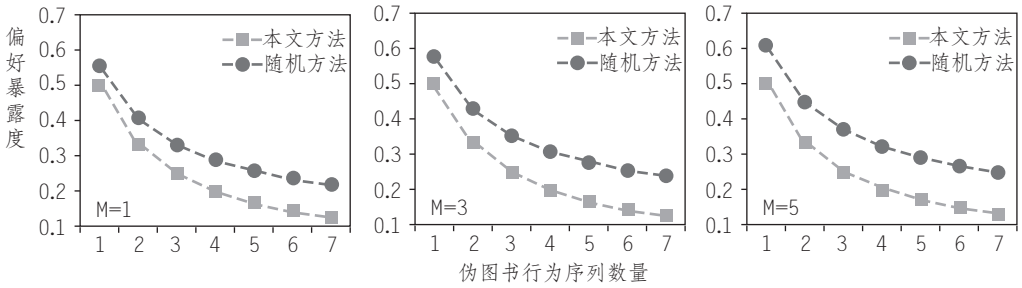

图 3　用户敏感偏好暴露程度评估

实验二旨在评估本文方法产生的伪行为序列与用户行为序列之间的特征相似度。这里使用“行为特征相似性”(参考定义 3.10 构建),以度量伪行为序列集$\mathbb{A}'=\{A_1,A_2,\cdots,A_m\}$关于用户行为序列 A_0 的特征相似性,即$\min_{A_k\in\mathbb{A}'}(\mathrm{sim}(A_k,A_0))$。显然,度量值越大越好,因为度量值越大意味着攻击者越难以通过特征分析,从行为序列集$\{A_0\}\cup\mathbb{A}'$中发现用户真实行为。该度量主要取决于用户行为序列长度和构造的伪行为序列数量。实验中,用户敏感偏好数量固定为 5。实验评估结果如图 4 所示,其中,子图左下角指示方法为每个用户序列所构造的伪行为序列数量(M=1、M=3 和 M=5)。可以看出,相比于随机方法,本文方法所生成的伪行为序列表现出更好的整体特征相似性。具体地,本文方法生成的伪行为序列与真行为序列之间的特征相似度接近于 1.0,即两者具有高度相似的特征(包括分布特征、连续特征和关联特征),并且即使在伪行为序列数量和长度发生改变的情况下,这种高度一致的相似性也几乎无改变。而随机方法生成的伪行为序列与真行为序列之间的整体特征相似性低于 0.15,明显低于本文方法。

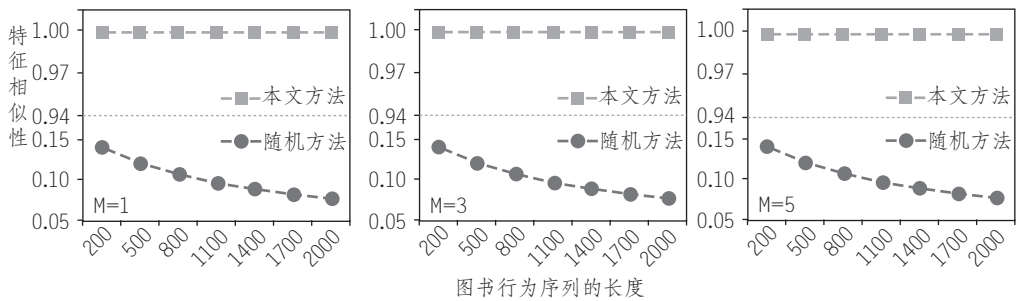

图 4　伪行为序列特征相似性评估

实验三采用人工方式测试本文方法生成的伪行为对用户真行为的混淆效果。实验挑选了 30 名本科生作为评价者,他们独立地从真假行为中挑选出最有可能的真行为。实验包括三个

部分。①第一组实验将同类行为划分为一系列独立的单个行为集(每个行为集中仅有一个真行为,其余均是算法为该真行为构造的伪行为),此时评价者只能根据行为本身特征来挑选可能的真行为。②第二组实验将同类行为划分为一系列独立的行为序列集(每个序列集仅有一个真行为序列,其余均是构造的伪序列),此时评价者能综合分析单个行为的分布特征和行为序列的连续特征,挑选可能的真行为。③第三组实验将图书浏览行为和阅读行为综合划分为一系列独立的行为序列集(每个序列既有浏览行为也有阅读行为),此时在挑选可能的真行为时,评价者可综合分析单个行为的分布特征、同类行为间的连续特征以及非同类行为间的关联特征。实验评估结果如图5所示,其中,子图左下角指示评价者考虑的行为特征,纵轴指示评价者成功找出用户行为的案例与该组实验所有案例的比率。可以看出,对于随机方法生成伪行为,即使作为混淆者的伪行为数量增加,评价者仍能容易从中找出用户真实行为。特别是在获知行为的分布特征、连续特征和关联特征后,评价者成功找出真行为的比率已经接近于1.0(即随机伪行为的混淆效果基本丧失)。相比于随机方法,本文方法所生成的伪行为表现出非常好的混淆效果:在只考虑单个行为的分布特征时,评价者找出真行为的概率与伪行为数量基本呈严格正反比关系;即使在综合考虑了行为分布特征、连续特征和关联特征时,评价者仍然难以从伪行为集中找出用户真实行为。综上,由于本文方法所生成的伪行为与用户行为具有很高的特征相似度,使得评价者难以根据特征分析找出用户行为,从而实现“真假难辨”的预期目标。

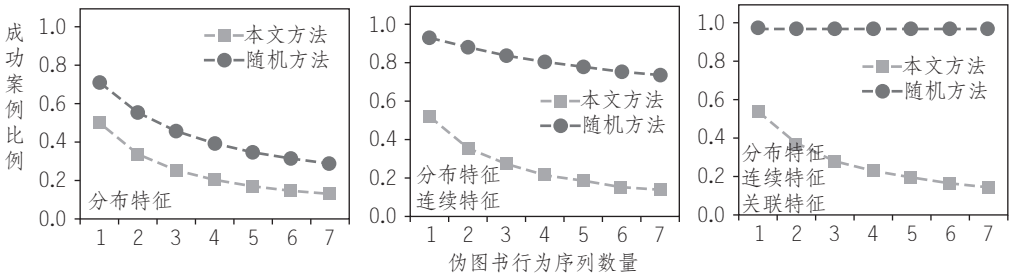

图5 伪行为序列混淆效果评估

5 分析讨论

从第1小节的框架图(见图1)可以看出,用户行为被混淆在一系列伪行为中,并以随机次序提交给服务器,但是由于来自同一个序列的各个行为之间具有很强的特征相关性,因此,借助于聚类等方法,攻击者仍能将服务器端所收集的图书服务请求记录划分为若干个独立的序列(即得到$A_0,A_1,A_2,\cdots,A_m$)。为此,我们假定攻击者获取了客户端所提交的全部图书服务请求(包括真服务请求和伪服务请求),并且已准确地将它们划分为独立的请求序列。此外,我们还假定服务器端的攻击者获取了运行在客户端的用户行为偏好保护算法副本。此时,攻击者能否根据掌握的服务请求序列集$\mathbb{A}'=\{A_0,A_1,A_2,\cdots,A_m\}$猜测出用户的任一敏感偏好$p^*\in\mathbb{P}^*$呢?以下,分三种情况分析讨论。

(1)在没有找出集合$\mathbb{A}'$中用户真实行为序列A_0的前提下,攻击者能否直接根据集合$\mathbb{A}'$猜测出用户敏感偏好p^*呢?此时,由于攻击者不知道集合$\mathbb{A}'$中哪个序列才是用户行为序列,他只能首先获取$\mathbb{A}'$各个行为序列相关的所有图书

偏好(可根据定义 3.1 计算得到),然后逐个去猜测这些偏好哪个是用户的敏感图书偏好。由于用户的任一敏感偏好 p^* 在 $\mathbb{A}'$ 中的暴露度相比于在 A_0 中的暴露度已经明显降低(见第 4 小节的实验评估),所以用户敏感偏好被猜测出来的概率将变得极小,变为原来的 $1/(m+1)$。换句话说,攻击者如果不找出用户真实行为序列 A_0,就难以猜测出用户敏感偏好 p^*。

(2)攻击者能否找出集合 $\mathbb{A}'$ 中用户真实行为序列 A_0 呢?此时,攻击者只能根据先验知识"用户真实行为序列会表现出富有规律的分布特征和连续特征"来猜测哪个才是用户真实行为序列。然而,由于本文方法所构造的伪行为序列与用户真实行为序列具有高度一致的可区分分布特征和连续特征,所以,攻击者难以根据行为序列的不同特征规律区分出真实行为序列。此外,由于攻击者掌握了丰富的背景知识,他还能根据行为序列 A_k 的各个子序列之间所表现出的关联特征来区分用户真实行为序列。但是在本文方法所构造的各个伪行为序列中,不同类别的子序列之间也具有与用户行为序列高度一致的关联特征,所以攻击者也难以根据行为序列的关联特征区分出真实行为。

(3)攻击者获取了运行在客户端的用户行为偏好保护算法的副本后,能否猜测出用户真实行为序列 A_0 呢?此时,攻击者可以首先将集合 $\mathbb{A}'$ 中的所有行为划分为若干个独立组,每组 A_k^* 均由 $(m+1)$ 个行为构成,记作:$A_k^* = \{a_k^0, a_k^1, a_k^2, \cdots, a_k^m\}$(其中,$a_k^t \in A_t, A_t \in \mathbb{A}'$)。然后,攻击者可以逐个输入 A_k^* 中的各个行为请求 a_k^t,然后观测用户行为偏好保护算法能否输出其余的行为请求 $A_k^* - \{a_k^t\}$。如果成功,则表明 a_k^t 是用户真实行为,进而推测出用户真实行为序列 A_0。然而,这样的尝试不会成功,因为在用户偏好保护算法中,每个伪行为请求均是从一个较大集中随机选取的(见算法 1 的步骤 3 和 13),即每次运行时,即使输入相同的数据也会输出不同的结果。

综上所述,虽然攻击者掌握着丰富的背景知识,但还是难以从服务端所记录的历史图书服务请求记录中识别出用户真实图书服务请求或者用户敏感个人图书偏好,因而本文方法具有较好的隐私安全性。结合前言和第 1 小节的内容可知:在安全性、准确性、高效性和可用性上,相比于已有方法,本文方法拥有更好的综合性能,能在不改变现有数字图书馆平台架构、不改变现有图书服务算法、不改变图书服务准确性、基本不改变图书服务效率的前提下,确保各类用户行为偏好隐私在不可信服务器端的安全性。

6 总结展望

本文提出了一个数字图书馆用户行为偏好隐私保护框架模型。该框架模型采用基于客户端的体系架构,通过在可信客户端为用户图书服务请求(即用户行为)精心构造一系列"真假难辨"的伪行为,"以假乱真"掩盖用户行为背后蕴含的敏感偏好。通过理论分析和评估实验,验证了该方法的有效性,即能在不损害数字图书馆服务的实用性、准确性和高效性的前提下,确保用户行为偏好隐私在不可信服务器端的安全性。然而,这项工作仍存在一些问题需要进一步研究改进。一、本文仅描述了一个较抽象的用户行为隐私保护框架模型。数字图书馆用户行为形式多样(如图书推荐行为、检索行为等),如何在模型框架下为各类用户行为设计实现相应的隐私保护算法还有待进一步深入研究。二、本文没有讨论数字图书馆用户行为隐私保护软件的具体设计实现问题。数字图书馆的用户终端界面形式多样(如移动应用终端、浏览器终端等),如何实现隐私保护软件与用户终端的无缝对接还有待进一步研究。

参考文献

[1] 易红，任竞. 图书馆大数据服务环境下用户隐私泄露容忍度的实证研究[J]. 图书馆论坛，2016，36（4）：57-64.（Yi Hong，Ren Jing. Empirical study on library users' privacy leakage tolerance under the background of library big data service[J]. Library Tribune，2016，36(4):57-64.）

[2] 彭华杰. 大数据时代图书馆读者的隐私危机与隐私保护[J]. 图书馆工作与研究，2014，1（12）：56-59.（Peng Huajie. The privacy crisis and protection of library readers in the big data era[J]. Library Work and Study，2014，1（12）：56-59.）

[3] 马晓亭，陈臣. 基于大数据生命周期理论的读者隐私风险管理与保护框架构建[J]. 图书馆，2016（128）：62-66.（Ma Xiaoting，Chen Chen. Construction of the privacy risk management and protection framework for library readers based on big data life cycle theory[J]. Library，2016（128）：62-66.）

[4] 宛玲，霍艳花，马守军. 英国大学图书馆网站个人信息保护政策文本分析及启示[J]. 图书情报工作，2016，60（12）：62-68.（Wan Ling，Huo Yanhua，Ma Shoujun. Text analysis of personal information protection policies of ten British university library websites and its enlightenment[J]. Library and Information Service，2016，60（12）：62-68.）

[5] Han Z，Huang S，Li H，et al. Risk assessment of digital library information security：a case study [J]. The Electronic Library，2016，34（3）：471-487.

[6] 邵志毅，杨波，梁启凡. 云计算中数字图书馆外包数据的完整性检测[J]. 图书馆论坛，2014（12）：98-103.（Shao Zhiyi，Yang Bo，Liang Qifan. Integrity verification of outsourced digital resources in cloud computing [J]. Library Tribune，2014（12）：98-103.）

[7] 王碧琴，任洁，冯彦平，等. 数字图书馆用户信息隐私的安全威胁分析[J]. 图书馆学研究，2015（10）：34-36.（Wang Biqin，Ren Jie，Feng Yanping，et al. Analysis on the safety threat of user's information privacy in digital library[J].Research on Library Science，2015（10）：34-36.）

[8] Wu Z，Xu G，Zong Y，et al. Executing SQL queries over encrypted character strings in the database-as-service model[J]. Knowledge-Based Systems，2012(35)：332-348.

[9] Wu Z，Xu G，Lu C，et al. An effective approach for the protection of privacy text data in the CloudDB[J/OL]. World Wide Web[2017-08-01]. https://link.springer.com/article/10.1007/s11280-017-0491-8.

[10] Wu Z，Shi J，Lu C，et al. Constructing plausible innocuous pseudo queries to protect user query intention[J]. Information Sciences，2015（325）：215-226.

[11] Wu Z，Li G，Liu Q，et al. Covering the sensitive subjects to protect personal privacy in personalized recommendation[J/OL]. IEEE Transactions on Services Computing[2016-10-08]. http://ieeexplore.ieee.org/document/7486070.

[12] 李东来，蔡冰，蒋永福，等. 以制度保障公共图书馆的读者权益[J]. 中国图书馆学报，2010，36（4）：17-23.（Li Donglai，Cai bing，Jiang Yongfu，et al.Protect reader's rights and interests in public library by rules and regulations[J]. Journal of Library Science in China，2010，36（4）：17-23.）

[13] 王琼，曹冉. 英国高校科研数据保存政策调查与分析[J]. 中国图书馆学报，2016，42（5）：102-115.（Wang Qiong，Cao Ran. Investigation and analysis of the data preservation policy in British universities[J]. Journal of Library Science in China，2016，42（5）：102-115.）

[14] 张雪梅，张艳芳，张阳. 图书馆读者隐私权的侵权表现与法律保护[J]. 情报科学，2012（6）：839-842.

(Zhang Xuemei, Zhang Yanfang, Zhang Yang. Tort performance and legal protection on reader privacy of library [J]. Information Science, 2012 (6): 839-842.)

[15] Mouratidis K, Yiu M. Shortest path computation with no information leakage[J]. VLDB Endowment, 2012, 5(8): 692-703.

[16] Khoshgozaran A H, Shirani-Mehr C. Blind evaluation of location based queries using space transformation to preserve location privacy[J]. Geoinformatica, 2013 (4): 599-634.

[17] 田丰, 桂小林, 张学军, 等. 基于兴趣点分布的外包空间数据隐私保护方法[J]. 计算机学报, 2014 (37): 123-138.(Tian Feng, Gui Xiaolin, Zhang Xuejun, et al. Privacy-preserving approach for outsourced spatial data based on POI distribution[J]. Chinese Journal of Computers, 2014 (37): 123-138.)

[18] Zhang F, Lee V E, Jin R. k-CoRating: filling up data to obtain privacy and utility[C]// Proceedings of the 20th AAAI Conference on Artificial Intelligence, 2014:320-327.

[19] XuY, Wang K, Zhang B. Privacy-enhancing personalized Web search[C]//Proceedings of the 16th International Conference on World Wide Web, 2007:591-600.

[20] Chen G, He B, Shou L, et al. UPS: efficient privacy protection in personalized Web search[C]//Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, 2011: 615-624.

[21] Shou L, Bai H, Chen K, et al. Supporting privacy protection in personalized Web search[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(2): 453-467.

[22] Narayanan A, Shmatikov V. Robust de-anonymization of large sparse datasets[C]//Proceedings of the IEEE Symposium on Security & Privacy, 2008:111-125.

[23] Juan V, Josep P, Juan H S. DocCloud: a document recommender system on cloud computing with plausible deniability[J]. Information Sciences, 2014, 258 (10): 387-402.

[24] Shang S, Hui Y, Hui P, et al. Beyond personalization and anonymity: towards a group-based recommender system[C]//Proceedings of the 29th Annual ACM Symposium on Applied Computing, 2014: 266-273.

[25] Josyula R R, Pankaj R. Can pseudonymity really guarantee privacy?[C]//Proceedings of the USENIX Security Symposium, 2000: 263-279.

吴宗大 温州大学瓯江学院温州市信息安全中心教授(特聘教授),硕士生导师。浙江 温州 325035。

谢 坚 温州市信息安全中心助理研究员。浙江 温州 325035。

郑城仁 温州大学数学与电子信息工程学院硕士研究生。浙江 温州 325035。

周志峰 温州大学图书馆信息部副主任,副研究员。浙江 温州 325035。

陈恩红 中国科学技术大学计算机科学与技术学院教授(国家杰青),博士生导师。安徽 合肥 230027。

(收稿日期:2017-10-03;修回日期:2017-12-10)