大数据环境下图书馆用户个人的信息保护研究

施航海[1],叶瑞哲[2],许卓斌[3]
(1 厦门理工学院 图书馆,福建 厦门 361024;2 厦门理工学院 信息中心,福建 厦门 361024;
3 厦门大学 信息与网络中心,福建 厦门 361005)

摘 要:本文从技术角度对大数据环境下图书馆用户个人的信息保护进行研究,探讨目前主流的用户个人信息保护技术,基于用户身份加密和属性加密的双重算法对用户信息保护的方法,结果满足图书馆个人信息保护需求,得出在大数据环境下用户身份和属性双重加密能够提高图书馆用户个人信息的保护能力.

关键词:大数据;图书馆用户个人信息;加密技术;加密算法
中图分类号:TP319 **文献标识码**:A **文章编号**:1000-7180(2018)05-0138-03
DOI:10.19304/j.cnki.issn1000-7180.2018.05.027

Research on the Personal Information Protection of Library Users Under Large Data Environment

Shi Hang-hai[1], Ye Rui-zhe[2], XU Zhuo-bin[3]
(1 Library, Xiamen University of Technology, Xiamen 361024, China;
2 Information Center, Xiamen University of Technology, Xiamen 361024, China;
3 Information and Network Center, Xiamen University, Xiamei 361005, China)

Abstract: The purpose of this paper is to study the protection of personal information for users of the Library under the environment of big data from a technical point of view, to explore the users' personal information protection technology, from the big data encryption algorithm based on the user environment the personal information of the proposed dual algorithm encryption and user identity attribute based encryption method to protect user information, the obtained results satisfy the library personal information protection requirements, can improve the protection ability of library users of personal information.

Key words: large data; personal information of library users; encryption technology; encryption algorithm

1 引言

随着大数据技术在图书馆中的应用,用户的阅读行为、身份信息、兴趣爱好等隐私特征被收集,并通过数据挖掘技术为用户提供个性化的图书定制服务,提高图书馆的服务水平.同时,用户的隐私成为图书馆重点保护的对象,不仅要加强用户信息管理的,更要提升图书馆用户个人信息保护的技术水平,采用先进的用户个人信息保护技术和复杂的数据加密算法是大数据环境下图书馆用户个人的信息保护的关键.

2 数据加密算法

2.1 对称加密算法

对称加密算法是加密与解密所使用的密钥相同的一种算法,当明文通过加密生成密钥后,使用相同的算法的逆算法进行解密,其算法公开、加密效率高.但是安全性相对较弱.常见的对称加密算法包括DES算法、RC算法、BlowFish算法等.DES算法将64位明文输入转换为64位密文输出,其中有8位作为奇偶校验,56位作为密码.RC算法是对初始数据簇进行随机搅乱,数据簇经过处理后得到多个不

收稿日期:2017-07-20;修回日期:2017-08-24
基金项目:福建省中青年教师教育科研项目(JZ170332);福建省中青年教师教育科研项目(JAT170436)

同的子密钥序列，再将子密钥序列与明文进行 XOR 运算，得到最后的密文. BlowFish 算法采用 64 位核心加密函数输出 64 位密文，使用一个 key 对原密钥进行变化得到新的密钥[1-5].

2.2　非对称加密算法

非对称加密算法是采用公开密钥和私有密钥进行数据加密和机密的算法，使用公开密钥加密的数据需要使用对应私有密钥进行解密，使用私有密钥加密的数据需要使用对应公开密钥进行解密. 常见的算法包括：RSA 算法和 Elgamal 算法等[6]. RSA 算法需要选择两个大的互异素数，素数 p 和 q，根据欧拉函数得到 $n=(p-1)(q-1)$，随机产生一个加密密钥 e，满足 e 与 n 互质，利用 Euclid 算法计算解密密钥 d，$de=1(\mathrm{mod}\varphi(n))$，$e$ 为公钥，d 为私钥. Elgamal 算法是通过计算有限域上离散对数实现加密，选择一个素数 P 和两个数随机数 g、x，计算公钥 $y=g\wedge x(\mathrm{mod}p)$，私钥为 x.

2.3　用户身份加密算法

用户身份加密算法是对用户身份信息进行加密的一类算法，算法包括 MD5 加密算法、ECC 椭圆曲线信息加密算法等[7]. MD5 加密算法是常见的细化散了算法，其对消息进行数据填充，再将消息以 512 位进行分组，一个分组进行 4 轮变化，输出 4 个变量，再通过 4 个变量进行下一分组的计算，直至最后一个分组的到的 4 个变量即为 MD5 值. ECC 椭圆曲线信息加密算法是利用特殊形式的椭圆曲线在有限域上构建方程 $y^2=x^3+ax+b(\mathrm{mod}p)$，式中 p 为素数；a、b 为正整数，a、b 小于 p，其满足条件 $4a^3+27b^2(\mathrm{mod}p)\neq 0$，点 (x,y) 和一个无穷点 o 组成椭圆曲线 E，计算 $Q=kP$ 的公式中，已知 Q、P 值求小于 p 的正整数 k 较为困难，而已知 k 和 P 计算 Q 却很容易，由此形成了 ECC 椭圆曲线信息加密算法.

2.4　属性加密算法

属性加密是身份密码体系的一个扩展，每一个用户所拥有的属性集合和所对应的密钥集都可以作为加密或者解密的条件，设一个集合 P 的每一个属性所获得的线性共享密钥部分能够构成一个 Z_p 上的向量，存在一个 l 行 n 列的分享生成矩阵 $\boldsymbol{M}$，$\boldsymbol{M}$ 的第 i 行为集合中的一个元素，使用函数 $\rho(i)$ 找到对应元素，向量 $v=(s,r_2,\cdots,r_n)$，s 为 Z_p 上的共享密钥，$r_2,\cdots,r_n$ 为随机数，则 Mv 得到的向量为 l 个元素所分享的信息[8].

3　大数据环境下图书馆用户个人的信息保护技术及算法应用

3.1　用户个人信息保护技术在图书馆中的应用

图书馆每天都产生海量的用户登记、查询、借阅等个人信息，这些信息大部分涉及个人隐私，为了保护图书馆读者用户的个人隐私，采用用户个人信息保护技术进行用户隐私保护必不可少. 首先要运用数据加密技术对读者进行电子信息的传输进行保护，利用匿名技术对用户在图书馆操作的信息进行保护. 其次在图书馆智能化物联网信息数据中，通过数据访问技术对传感网中的用户信息进行保护. 最后，在向用户进行用户信息推荐时，根据读者历史的浏览记录、阅读记录和评论记录等进行数据采集，并进行记录信息挖掘，构建用户兴趣模型，进而能够为用户定制个性化服务. 采用差分保护技术对用户隐私进行保护.

在大数据环境下，图书馆用户信息推荐过程中，用户与图书之间的关系是最基本的数据，利用评分矩阵对用户阅读行为进行分析，建立二者之间的关系可以设有 n 个用户对 m 个用户信息的评分矩阵 $\boldsymbol{R}_{n\times m}$，其中 r_{ui} 为用户 u 对用户信息 i 的评分. 用户因素矩阵可以表示为 $\boldsymbol{P}_{n\times d}$，项目因素矩阵可以表示为 $\boldsymbol{Q}_{d\times m}$，假设 $\bar{R}_n\times m=\boldsymbol{P}_{n\times d}\times\boldsymbol{Q}_{d\times m}$. 采用差分技术进行个人隐私保护是要将其引入到图书关联的矩阵中，在原函数输出值上加上 $Lap(\frac{\Delta f}{\in})$，就得每一个用户信息的平均分，$IAvg(J)=\frac{\sum_{Rj} r_{ui}}{IR_{j^1}}$，式中，$j$ 表示第 j 个用户信息；$|R_j|$ 表示参与第 j 个用户信息评分的用户数. 评分的敏感度 $\Delta r=r_{\max}-r_{\min}$.

3.2　数据加密算法在图书馆中的应用

在大数据环境下图书馆用户个人信息保护中，本文采用用户身份加密算法与属性加密算法相结合方法确保图书馆用户个人信息安全. 通过一个密钥抽取算法对集合属性(ω') 生成对应非线性密钥，再对明文 M 进行加密，解密时通过所拥有的属性私钥集合 ω 进行解密，满足条件 $(\omega'\cap\omega)>d$，式中 d 为系统设定值. 设 G_1 为一个以素数 p 为阶的双线性群，g 为生成元，双线性配对计算 $e:G_1\times G_1\rightarrow G_2$.

计算拉格朗日参数：

$$\Delta_{i,s}(X)=\prod_{j\in S,j\neq i}\frac{x-j}{i-j} \tag{1}$$

式中 S 为 Z_p 上的一个集合，其属性集为 U.

加密与解密的过程包括初始设置、密钥抽取、加密计算和解密计算四个步骤.

初始设置在系统用户身份加密中,公钥设为 $Y=e(g,g)^{y}$,管理密钥设为$\{t_1,\cdots,t_{|U|}\}$,y.

属性加密算法在一个集合属性随机选择 Z_p 上的 t_i 作为私钥,其对应的公钥为:$\{T_1=g^{t1},\ldots,T_{|U|}=g^{t|U|}\}$.

密钥抽取是对一组属性集合 $\omega\subseteq U$ 随机选择一个 $d-1$ 维的多项式 $q(x)$,$q(0)=y$,用户的私钥为 D_i,ω 对应 T_i 满足 $D_i=g^{\frac{q(i)}{t_i}}$.

加密计算对集合 ω' 和明文 $M\in G_2$ 选择随机数 s 得到 E:

$$E=(\omega',E=MY^{s},\{E_i=T_i^{s}\}_{iE\omega'}) \tag{2}$$

解密计算对集合 ω 满足$(\omega'\cap\omega)>d$ 的条件下任意选择两个属性的交集,利用公式(4)得到:

$$E/\sum_{i\in S}(e(D_i,E_i))^{\Delta is(0)}=M \tag{3}$$

该算法将身份加密算法与属性加密算法相结合,可以将图书馆用户 ID 分解出性别、年龄、借阅时间、工作单位、图书类型等多个集合,进而对用户身份的访问和属性信息的认证进行加密.

4 结束语

本文对大数据环境下的图书馆用户个人信息保护相关技术进行研究,现代化的图书馆不仅是图书借阅和浏览的场所,更是各类信息汇聚的场所,其中个人信息的安全是图书馆信息管理工作的主要内容.在大数据环境下,图书馆不仅实现了自动化的借阅和信息服务,而且能够针对用户个人信息进行分析为其推荐个性化的服务,为此在用户个人信息保护技术方面,本文对数据加密技术、匿名技术、差分保护技术和数据访问控制技术进行了阐述,探究技术的原理,并对常见的数据加密算法进行介绍,将各项技术计算法交叉应用到图书馆用户个人的信息保护中提高图书馆个人隐私的安全保护能力,为促进图书馆用户个人信息保护技术发展提供借鉴参考.

参考文献:

[1] 董永为.数据加密技术在计算机网络通信安全中的应用分析[J].网络安全技术与应用,2016(4):39-40.

[2] 莫家庆,胡忠望,林瑜华.基于可信计算的匿名通信系统方案研究[J].计算机应用与软件,2016,33(12):84-88.

[3] 高翰卿,秦小麟,史文浩.基于目的和上下文推理的数据库访问控制模型[J].计算机科学与探索,2016,10(9):1229-1239.

[4] 鲜征征,李启良.差分隐私保护在推荐系统中的应用研究[J].计算机应用研究,2016,33(5):1549-1553.

[5] 张文文,炳勋.基于 RSA 与 DES 的多重加密可信加密算法[J].电脑迷,2016(9):33-35.

[6] 车念,赵士元,丁莎.融合多重加解密算法的保密通信系统[J].计算机工程与设计,2017,38(4):936-940.

[7] 李新,彭长根,牛翠翠.隐藏树型访问结构的属性加密方案[J].密码学报,2016,3(5):471-479.

[8] 韩礼红,韩翠峰.大数据时代图书馆个性化信息服务中读者隐私保护研究[J].阜阳师范学院学报:社会科学版,2016(1):153-156.

作者简介:

施航海 男,(1974-),馆员.研究方向为数字图书馆技术、信息系统建设.E-mail:SHH@xmut.edu.cn.

叶瑞哲 男,(1976-),工程师.研究方向为校园信息化建设与管理.

许卓斌 男,(1975-),硕士,工程师.研究方向为数据中心、园区网、信息化应用建设、并行计算、大数据分析等.