

基于区块链的可信高校智慧图书馆研究*

杜 兰,陈琳琳,张 丽,田 越

(南京理工大学紫金学院,江苏 南京 210023)

摘 要: 云计算、大数据、区块链和人工智能等新技术为图书馆的快速发展带来了新的机遇,很多高校图书馆开始尝试将人脸识别技术和推荐系统引入智慧图书馆建设中,但是这两门技术都面临着隐私泄露和可信问题。区块链是去中心化、公开可信且不可篡改、可追溯、可验证、可审计、不可抵赖的新颖的分布式账本,区块链技术特性很适宜实现隐私保护。为此提出在区块链下进行可信高校智慧图书馆建设研究,并用 Python 语言编程验证了区块链、人脸识别和推荐系统融合的可能性,为高校智慧图书馆建设提供一种较新的思路。

关键词: 区块链;智慧图书馆;人脸识别;推荐系统;隐私保护

中图分类号: TP309.2

文献标识码: A

DOI: 10.19358/j.issn.2096-5133.2020.12.002

引用格式: 杜兰,陈琳琳,张丽,等. 基于区块链的可信高校智慧图书馆研究[J]. 信息技术与网络安全, 2020, 39(12): 7-12.

Research on trusted university smart library based on blockchain

Du Lan, Chen Linlin, Zhang Li, Tian Yue

(Nanjing University of Science and Technology ZiJin College, Nanjing 210023, China)

Abstract: New technologies such as cloud computing, big data, blockchain and artificial intelligence have brought new opportunities for the rapid development of libraries. Many university libraries try to introduce face recognition technology and recommendation system into the construction of smart library. But both have privacy and credibility problems. Blockchain is a decentralized, publicly trusted and non-tampering, traceable, verifiable, auditable and non-repudiation novel distributed ledger. The features of blockchain technology are suitable for privacy protection. Therefore, this paper puts forward the research on the construction of trusted university smart library under the blockchain. In this paper, the possibility of integrating blockchain, face recognition and recommendation system is verified by Python programming, which provides a new idea for university smart libraries.

Key words: blockchain; smart library; face recognition; recommendation system; privacy protection

0 引言

如今,大数据、区块链和人工智能等技术席卷各个领域,冲击了传统图书馆的运行模式,驱动着图书馆向数字化、智能化和智慧化演进。智慧图书馆的核心是“感知、智能、融合”,它借助高新智能技术来推动图书馆转型,满足读者个性化、专业化和人性化的智慧型知识服务需求^[1]。智慧图书馆服务实践主要有基于 RFID/手机/网络的自助借还服务、智能清点/定位服务、智能座位预约服务、

3D/AR/VR 虚拟导航服务等^[2]。

近年来,人脸识别技术在人工智能生物特征识别领域发展迅速,已在大众生活中被广泛使用。然而,人脸数据泄露事故频发,人脸识别技术的安全性与隐私保护受到广泛关注。人脸识别技术并不是本身不安全,而是人脸数据没有有效监管从而存在隐私泄露的风险,因此研究在安全可信的环境下进行人脸识别迫在眉睫。与此同时,推荐系统已逐渐成为智慧图书馆研究人员关注的研究热点。然而推

* 基金项目: 江苏省高校自然科学基金项目(19KJB520039); 江苏省高校哲学社会科学研究一般项目(2019SJA2056); 南京理工大学紫金学院 2018 年度科研项目(2018ZRKX0401001)

荐系统的基石是用户历史数据集,它模糊了隐私和隐私数据的边界,存在隐私泄露风险。因此,在有效保护用户隐私的基础上进行精准推荐是一项值得挑战的工作。

在不可信的开放“互联网+”时代,区块链恰逢其时,它的技术特性使其很适合存储和保护隐私数据,给人脸识别技术和推荐系统的隐私保护带来发展契机。因此,本文设计与实现了一个基于区块链的可信智慧图书馆系统,充分运用人脸识别技术和推荐系统的优势,为用户提供方便。

1 相关研究

1.1 区块链及其在图书馆领域的研究及应用

1.1.1 区块链概述

区块链是整合了分布式存储、P2P 传输服务、共识机制、加密算法等技术的分布式记账技术,它通过加密块链式区块结构来生成和更新数据,利用自动化脚本代码生成的智能合约来编程和操作数据。它具有去中心化(Decentralized)、免信任(Trustless)、时间戳(Time Stamp)、加密算法(Cryptography)和智能合约(Smart Contract)等技术特征。

1.1.2 区块链在图书馆领域的研究与应用

国内外图书馆界已经开展了很多相关研究工作。ROD C 在 2016 年指出区块链是数字资产的新架构,开启了区块链技术在图情档领域的研究^[3]。2018 年 4 月美国图书馆协会(American Library Association, ALA)将区块链技术列为未来图书馆重大变革性技术。MANLEY S 在 2019 年提出区块链有助于保护版权^[4]。区块链在我国图书馆的应用研究有信息安全共享模型^[5]、读者隐私保护^[6]、数字版权管理^[7]等。

1.2 人脸识别及其在图书馆领域的研究与应用

1.2.1 人脸识别概述

自从 1988 年人脸身份识别首篇论文^[8]发表以来,经过几十年的发展,人脸识别技术已经日臻成熟。目前,人脸识别技术已广泛应用在支付系统、门禁系统、身份验证等业务上。

1.2.2 人脸识别在图书馆领域的研究与应用

国内外已有很多关于人脸识别在图书馆领域的相关研究,RAVIKANTH M 等设计与实现了新的图书馆门禁系统^[9];周东等介绍了中国计量大学图书馆开发人脸识别应用系统的实施方案^[10];西南大

学、哈尔滨工业大学和上海海事大学等开展了自助刷脸借书和刷脸预约座位等应用。

1.3 推荐系统在图书馆领域的研究及应用

1.3.1 推荐系统概述

推荐系统是解决信息过载的利器,目前已被 Google、Amazon、百度、淘宝、去哪儿网、携程网实际应用,也得到了良好的推荐效果。

1.3.2 推荐系统在图书馆领域的研究及应用

推荐系统在我国图书馆的研究主要有基于聚类的推荐方法、基于协同过滤的推荐方法、基于关联规则的推荐方法、混合推荐方法和基于用户画像的推荐方法^[11]。

2 区块链下可信智慧图书馆系统的总体设计

2.1 系统流程

系统流程图如图 1 所示。

(1)用户注册。将用户注册信息(姓名、照片、年龄、性别、专业、职业、地址等)录入数据库。

(2)入馆。用户 A、用户 B 刷脸进行区块链和人脸识别双重身份认证。

(3)个性化推荐。系统根据区用户最近的行为信息、内容信息和互动信息推荐相关书籍。

(4)借/还书。用户 A 提出借书申请,如果图书馆里有可借阅的书籍,则用户 A 既可以从图书馆直接借走,也可以通知目前图书馆区块链上记录的这本书的借阅者用户 B 直接借给用户 A。

用户 B 提出还书申请,既可以直接还给图书馆,也可以直接借给经图书馆区块链上记录的图书借阅申请者用户 A。

(5)互动。用户提出互动申请,推荐引擎会把学习交流、讨论分享信息推荐到具有相同兴趣的其他用户。

2.2 系统架构

本文基于 Django 开源框架,使用 Fabric 的 Python SDK 与区块链网络通信来搭建去中心化的可信智慧图书馆系统,其整体架构图如图 2 所示。

(1)视图层:系统用户交互层,是使用 HTML5 编写的系统各种应用程序页面。

(2)业务逻辑层:系统的核心业务层,该层的基础是使用 Hyperledger Composer 的一系列组件工具向上提供 REST APIs 服务接口。

(3)数据资源层:MySQL 存储系统中的基础数据,Fabric 将数据地址保存在区块链中。

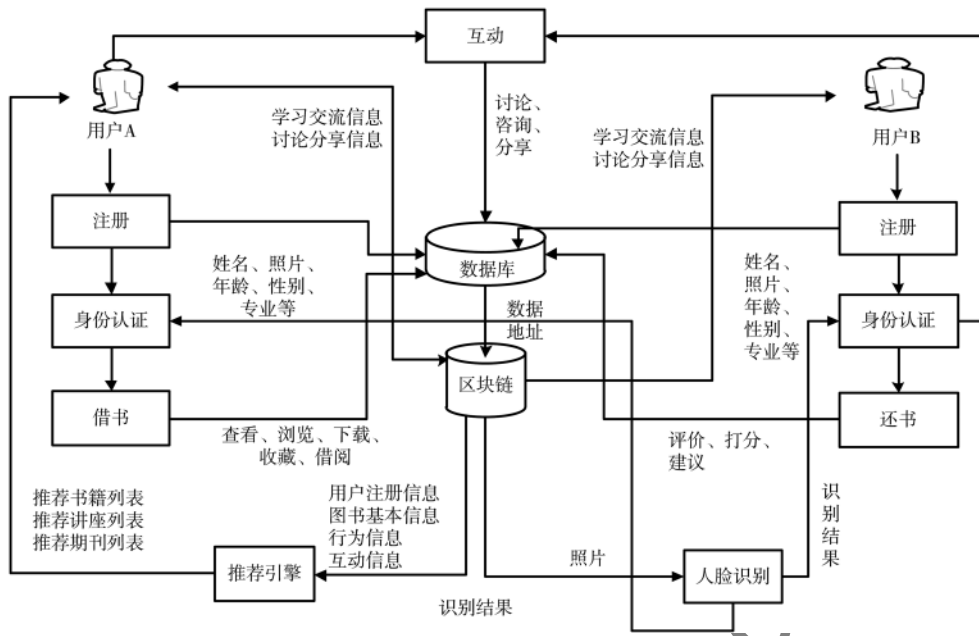


图1 系统流程图

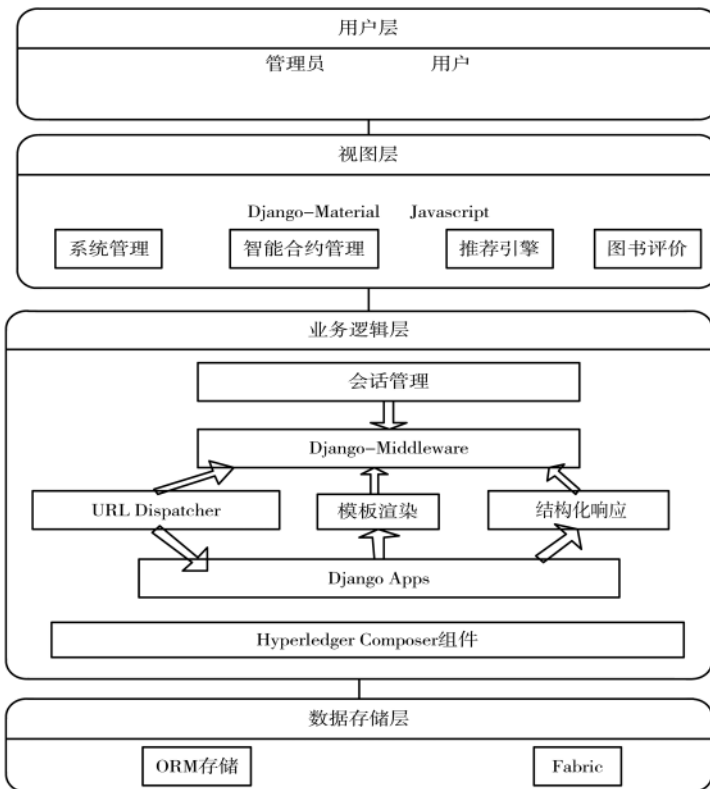


图2 系统架构图

程如图3所示。

(1)人脸采集:当用户首次登录系统时,需要上传一张真实的人脸照片来完成注册操作。注册成功后,系统返回Public Key。系统获取上传的图像,经过预处理、提取人脸特征值、对人脸特征降维之后,形成人脸特征数据集,再进行人脸特征分类训练后将训练好的人脸信息数据写入数据库,将数据地址和数据操作写入区块链并验证。

(2)数据匹配与处理:在系统进行身份认证时,利用图书馆的摄像头实时采集用户的人脸视频提取图像,提取人脸特征形成数据集,并根据用户ID从区块链数据库中获取数据集的数据库原始地址和操作记录,查找到该用户的注册特征数据集,两者比对,并返回身份认证结果。

(3)权限应用:人脸识别成功,身份认证通过以后,系统便会分配对应的系统权限和新Public Key,用户便可无障碍地享受图书馆的多种服务,比如门禁考勤系统实现人脸识别、统计用户进馆时长等功能;图书借阅管

理系统实现图书自助借阅、归还和续借等功能;座位管理系统实现自助预约选座、预约查询、座位使用查询等功能;讨论室预约系统实现讨论室预约查询和预订功能;自助服务设备系统实现自助打印复

3 关键技术分析

3.1 区块链与人脸识别

3.1.1 总体设计

结合人脸识别技术的图书馆双重身份认证流

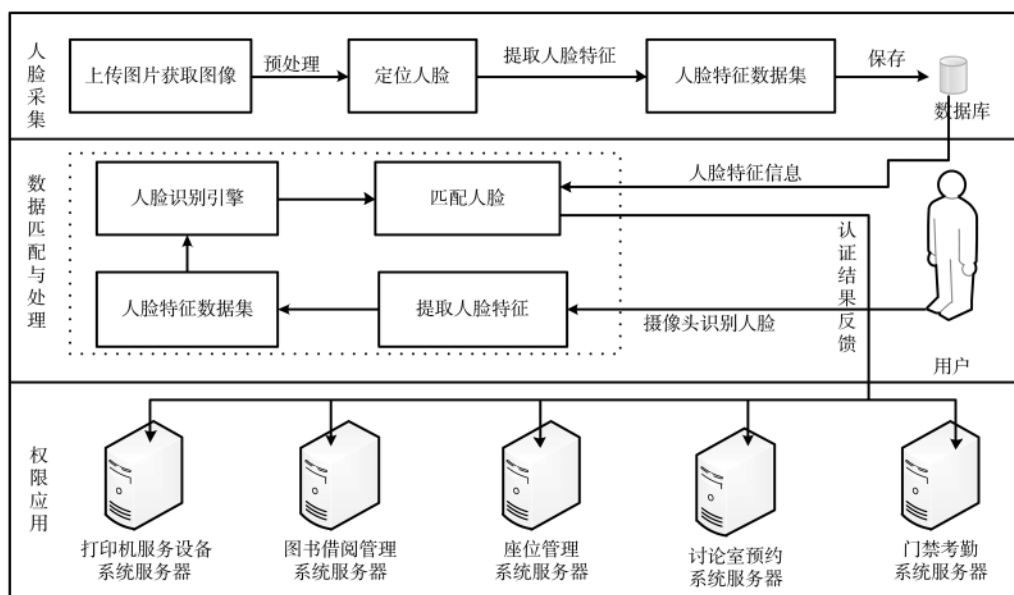


图3 区块链和人脸识别双重身份认证流程

印扫描、自助上网等功能。

3.1.2 主要算法

(1)主成分分析(PCA)算法理论

PCA 的原理可以表述为:矩阵的主成分就是其协方差对应特征向量,按照特征值的大小进行排序,最大的特征值就是第一主成分,然后以此类推。PCA 从 n 维减少到 k 维的算法描述如下:

输入:数据集 X 的 m 个 n 维特征值, $X=[x_1, x_2, \dots, x_n]$

①均值归一化。计算 X 所有特征的均值 μ_j , 然后令 $x_j = x_j - \mu_j$ 。如果特征是在不同的数量级上,还需要将其除以标准差 σ^2 。

②计算协方差矩阵 (covariance matrix), 令 $U = \frac{1}{m} XX^T$ 。

③计算协方差矩阵 U 的特征向量(eigenvectors)和特征值。

④将特征向量按特征值大小排列成矩阵,取前 k 个向量,获得 $n \times k$ 的矩阵 U_{reduce} 。

输出:新特征向量 $Y = (U_{\text{reduce}})^T X$, Y 就是 X 从 n 维减少到 k 维的数据。

(2)支持向量机(SVM)算法理论

SVM 是一种监督学习方法,主要目的是对数据进行二元分类,最终在一个二分类问题中找出最优的决策面。非线性支持向量机学习算法步骤如下:

输入: $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, x_i \in \mathbf{R}^n, y_i \in \{+1, -1\}, i \in (1, 2, \dots, n)$

①选取适当的核函数 $k(x, z)$ 和惩罚参数 C , 构造并求解凸二次规划问题:

$$\min_{\omega, b} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^n a_i$$

$$\text{s.t.} \sum_{i=1}^n a_i y_i = 0$$

$$0 \leq a_i \leq C, i = 1, 2, \dots, n$$

得到最优解 $a^* = (a_1^*, a_2^*, \dots, a_n^*)^T$ 。

$$\text{②计算 } \omega^* = \sum_{i=1}^n a_i^* y_i x_i$$

选择 a^* 的一个分量 a_i^* 满足条件 $0 < a_i^* < C$, 计算

$$b^* = y_j - \sum_{i=1}^n a_i^* y_i K(x_i \cdot x_j)$$

③求分类决策函数: $f(x) = \text{sign}(\sum_{i=1}^n a_i^* y_i K(x_i \cdot x_j) + b^*)$

如选取高斯核函数 $K(x, z) = \exp(-\frac{\|x-z\|^2}{2\sigma^2})$,

在此情况下,分类决策函数为 $f(x) = \text{sign}(\sum_{i=1}^n a_i^* y_i \exp(-\frac{\|x-z\|^2}{2\sigma^2}) + b^*)$ 。

输出:分类决策函数。

3.2 推荐引擎

3.2.1 总体设计

结合协同过滤推荐算法的推荐引擎如图 4 所示。

(1)数据采集处理:数据采集处理层采集到用户的原始数据信息后,需要进行数据挖掘和数据预处理,得到处理后数据,提取用户的主要特征,形成用户特征库,保存到用户偏好数据库中。

(2)个性化推荐:用户请求推荐,推荐引擎启动协同过滤算法,匹配目标最近邻居集,最终得到推荐结果。

(3)个性化服务层:个性化推荐结果为用户提供个性化服务,包括个性化检索、个性化阅读推荐、可视化展示、交流互动分享、评价反馈等模块。

3.2.2 主要算法步骤

基于用户的协同过滤算法分为 3 个步骤:

(1)计算用户之间的相似度;(2)利用统计学方法进行邻居选择;(3)根据近邻的历史行为计算预测值。该算法关键问题是计算用户之间的相似度,本文使用皮尔逊相关系数计算相似度,如式(1)所示:

$$\text{sim}(i, j) = \frac{\sum_{k \in I_i \cap I_j} (r_{ik} - \bar{r}_i)(r_{jk} - \bar{r}_j)}{\sqrt{\sum_{k \in I_i \cap I_j} (r_{ik} - \bar{r}_i)^2} \sqrt{\sum_{k \in I_i \cap I_j} (r_{jk} - \bar{r}_j)^2}} \quad (1)$$

其中, $I_i \cap I_j$ 是用户 i 和用户 j 阅读记录的交集。 r_{ik} 和 r_{jk} 分别表示用户 i 和用户 j 对交集集中的每个具体书目 k 的评分, $\bar{r}_i$ 和 $\bar{r}_j$ 表示用户 i 和 j 交集中所有书目的评分的平均值。

4 结论

随着区块链和人工智能技术的日趋成熟,把这些技术应用在现实场景的研究将会越来越多。本文设计与实现的区块链下的可信高校智慧图书馆系统,首先利用区块链保障图书馆信息资源可信共享,其次,基于人脸识别和区块链实现双重身份认证,最后,设计协同过滤推荐系统提供用户个性化服务。本系统使用 Django+Python+MySQL 技术路线来编程实现并进行验证,系统的响应速度在 150 ms 以内,并发量达到了 500,结果表明该系统能够快速对较多用户的同时操作进行响应,满足图书馆系统需求。但是在具体实际推广中还有很多可完善的地方,主要体现在 Fabric 区块链目前只支持数据存

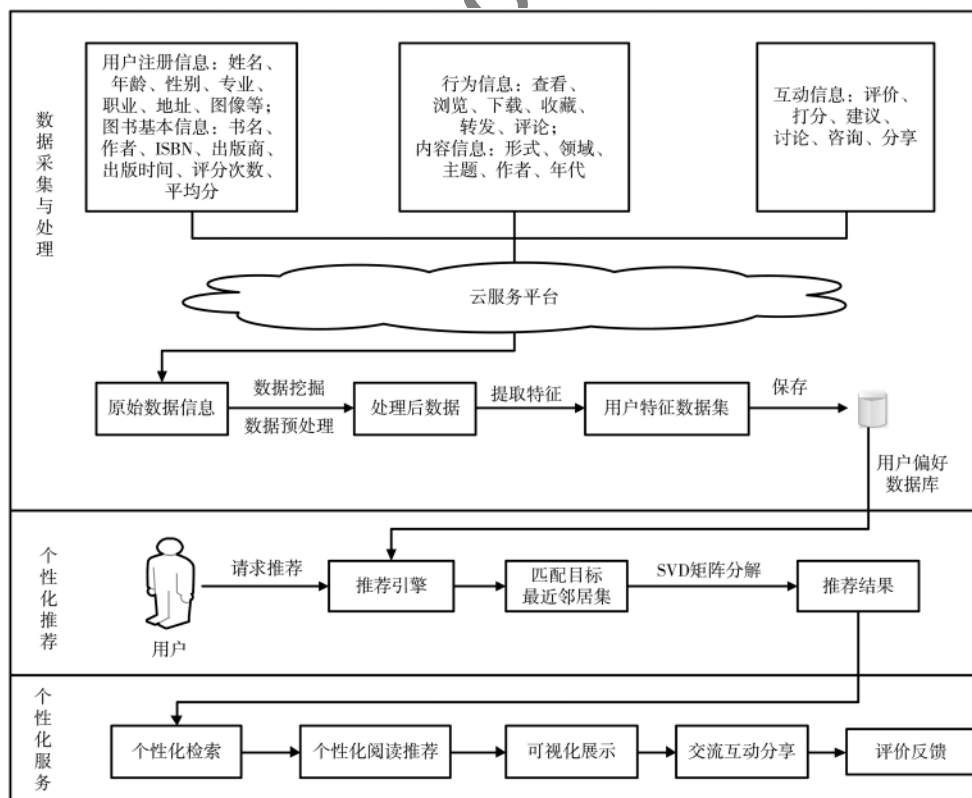


图 4 结合协同过滤推荐算法的推荐引擎

与查询,功能稍显单薄,而如何克服区块链在数据搜索方面速度慢的问题,提高链上查询效率也是下一步的研究方向。

参考文献

- [1] 王世伟.未来图书馆的新模式:智慧图书馆[J].图书馆建设,2011(12):1-5.
- [2] 陆婷婷.从智慧图书馆到智能图书馆:人工智能时代图书馆发展的转向[J].图书与情报,2017(3):98-101,140.
- [3] ROD C.Blockchain:a new architecture for digital content[J].ECONTENT,2016,39(8):22-23.
- [4] MANLEY S.On the limitations of recent lawsuits against Sci-Hub, OMICS, researchgate, and georgia state university[J].Learned Publishing,2019,32(4):375-381.
- [5] 徐俐华,聂飞霞.基于数字图书馆联盟链的信息资源安全共享模型构建[J].图书情报工作,2020(3):53-58.
- [6] 柳林子,赵力.区块链技术下图书馆读者个人信息保护研究[J].图书馆工作与研究,2019(5):96-101.
- [7] 秦珂.区块链技术视野下的图书馆数字版权管理:作用机制、创新价值和建设[J].图书馆论坛,2020(4):113-122.
- [8] GALTON F.Personal identification and description[J].Nature,1888,38:173-177.
- [9] RAVIKANTH M,HITESH B,KRISHNA K B.Library attendance for students with face recognition[J].International Journal of Advance Research and Innovative Ideas in Education,2018(4):1346-1349.
- [10] 周东,施芒.图书馆自行开发人脸识别应用系统的实践与思考——以中国计量大学图书馆为例[J].图书馆工作与研究,2020(1):80-87.
- [11] 张晗,毕强,李洁,等.基于用户画像的数字图书馆精准推荐服务体系构建研究[J].情报理论与实践,2019,42(11):69-74,51.

(收稿日期:2020-08-25)

作者简介:

杜兰(1986-),女,硕士,高级工程师,主要研究方向:区块链、人工智能。

(上接第6页)

- 01-xx)[2020-06-16].<http://blog.nsfocus.net/wp-content/uploads/2020/01/2019-Cybersecurity-Insights.pdf>.
- [2] 冯登国,张敏,李昊.大数据安全与隐私保护[J].计算机学报,2014,37(1):250-262.
- [3] DWORKIN M.Recommendation for block cipher modes of operation:methods for format-preserving encryption[M].Gaithersburg:NIST Special Publication,2016.
- [4] DWORK C.Differential privacy-Encyclopedia of cryptography and security[M].Boston,MA:Springer,2011.
- [5] SWEENEY L.K-anonymity:a model for protecting privacy[J].International Journal of Uncertainty Fuzziness & Knowledge Based Systems,2002,10(5):557-570.
- [6] MACHANAVAJJHALA A,KIFER D,GEHRKE J,et al.L-diversity:privacy beyond K-anonymity[J].ACM Transactions on Knowledge Discovery from Data,2007,1(1):1-52.
- [7] ISO/IEC.Privacy enhancing data de-identification terminology and classification of techniques:20889[S].2018.
- [8] Article 29 Data Protection Working Party.Opinion 05/2014 on anonymisation techniques[R/OL].(2014-03-xx)[2020-06-16]./<https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/>.
- [9] 全国信息安全标准化技术委员会.信息安全技术—信息安全风险评估规范:20984[S].2017.
- [10] SHANNON C E.A mathematical theory of communication[J].The Bell System Technical Journal,1948,27(3):379-423.

(收稿日期:2020-10-20)

作者简介:

陈磊(1991-),男,博士,绿盟和清华大学联合培养博士后,主要研究方向:数据安全、隐私保护、区块链。

薛见新(1984-),男,博士,绿盟和清华大学联合培养博士后,主要研究方向:图计算、网络攻击溯源、机器学习。

刘文懋(1983-),通信作者,男,博士,高级工程师,主要研究方向:网络安全、云安全、安全机器学习。E-mail:liuwenmao@nsfocus.com。

版权声明

经作者授权，本论文版权和信息网络传播权归属于《信息技术与网络安全》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《信息技术与网络安全》编辑部
中国电子信息产业集团有限公司第六研究所