文章编号：1003-1421(2020)11-0080-06 中图分类号：U29-39 文献标识码：A
DOI：10.16668/j.cnki.issn.1003-1421.2020.11.14

基于区块链技术的铁路数据汇集共享体系架构研究

Architecture of the Railway Data Collection and Sharing System Based on Blockchain

代明睿
DAI Mingrui

（中国铁道科学研究院集团有限公司 电子计算技术研究所，北京 100081）

(Institute of Computing Technologies, China Academy of Railway Sciences Corporation Limited, Beijing 100081, China)

摘 要：伴随铁路智能化建设，我国铁路已经面向建造、装备、运营等板块相关应用系统进行全业务、全类型的数据汇集，其中敏感数据信息的隐私保护和安全可控需要重点关注。从铁路数据汇集共享现状出发，分析基于区块链技术的铁路数据汇集共享，设计基于区块链技术的铁路数据汇集共享体系架构，研究基于区块链的铁路数据汇集共享关键技术，即基于区块链的铁路数据汇集技术、数据审计技术、数据检索技术和数据共享技术，分析铁路数据汇集共享体系架构安全特性。研究能够同时兼顾数据全面汇集和数据拥有方自主可控的需求。

关键词：铁路；区块链；星际文件系统；铁路数据汇集；共享体系；架构；数据汇集；数据审计；数据检索；安全性

Abstract: With the construction of intelligent railway in China, data collection covers all business and all types from application systems related to construction, equipment and operation. More attention should be paid to privacy protection and security of sensitive data and information. After analyzing the current situation of railway data collection and sharing, this paper proposes a railway data collection and sharing system, combining blockchain technology, studies the key technologies of railway data collection and sharing, namely railway data collection, data audit, data retrieval and data sharing technology based on blockchain, , and analyzed the security and characteristics of railway data collection and sharing system. This research has simultaneously taken into account the comprehensive data collection and independently controllable needs of the data owner.

Keywords: Railway; Blockchain, IPFS, Railway Data Collection; Sharing System; Architecture; Data Collection; Data Audit; Data Retrieval; Security

1　基于区块链技术的铁路数据汇集共享体系

1.1　基于区块链技术的铁路数据汇集共享

铁路数据汇集共享是将铁路在运输生产和经营活动中产生和积累的海量数据进行汇集存储并在各业务间进行共享利用的过程。数据汇集共享是围绕铁路运输生产、建设管理、经营开发战略决策等领域开展铁路数据综合分析的必要条件，有利于发掘和释放数据资源的潜在价值，发挥数据在提高效率、提升效益、确保安全、优化服务等方面的独特作用，提升铁路管理水平和服务质量，增强安全风险防控能力，实现铁路运营和管理的精准性、高效性、预见性。目前，遵循数据“一点采集、全程应用”的原则，铁路主要信息系统业务数据向数据中心的汇集工作已经展开。数据中心的铁路数据汇集以铁路基础数据管理和数据整合共享为抓手，能够盘活铁路数据资产，充分挖掘铁路数据价值，有效支撑铁路各专业开展数据分析、成果共享和产品研发，有力推动铁路信息系统的互联互通、数据共享和深度集成。

区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式，其本质是一个分布式公共账本，区块链系统中的参与者可以共同维护账本，但单一用户不可以对它进行控制。数据的完整、完全与访问控制管理是数据存储与共享中的关键问题，目前铁路数据汇集采用传统集中存储方式实现，这也带来了多方面的问题。第一，集中存储、管理铁路领域海量业务数据将面临多方面的安全风险，网络攻击时可能导致数据泄露，因而如何保证数据安全是需要考虑的关键问题。第二，传统数据中心的存储管理机制导致汇集后数据的提供者失去对数据的自主控制权，这种数据的不可控性既带来了安全隐患，也增加了核心敏感数据汇集的难度。第三，传统中心化架构的访问控制管理权限划分粒度低，难以实现对数据共享的精细管控，虽然针对这类问题近年来提出了一些解决方案，但技术复杂且实现成本较高。

区块链作为一种新兴技术，具有去中心化、难以篡改、共识信任等特点，被广泛用于金融、物流等领域[1-2]。星际文件系统 (IPFS) 是一个点对点的分布式版本文件系统，基于分布式哈希表解决数据的传输、定位和存储问题，数据保存到星际文件系统节点后，系统将返回基于该信息计算得出的唯一哈希值。哈希值与数据内容一一对应保证信息不被篡改。区块链与星际文件系统的特点相结合[3-6]，提供了解决铁路数据汇集面临各类问题的可能。因此，以区块链技术为基础，针对铁路数据共享的业务特点提出一种安全、高效的数据汇集共享体系架构，研究其中的关键技术问题。业务信息系统可以在不泄露数据明文的情况下完成数据汇集，将索引信息上链存储，数据的使用需要得到业务方的授权和管控，多重加密保证数据真实安全，共识机制保证数据存管用全程留痕。该架构可以实现数据全面汇集和数据自主可控的结合。

1.2　基于区块链技术的铁路数据汇集共享体系架构

结合现阶段铁路数据汇集共享的实际需求，基于区块链技术的特征与优势，设计基于区块链技术的铁路数据汇集共享体系架构。基于区块链技术的铁路数据汇集共享体系架构如图 1 所示。

由图 1 可知，基于区块链技术的铁路数据汇集共享体系架构主要包含数据存储层、区块层、合约层、汇集共享应用层和服务层 5 部分。

（1）数据存储层。区块链并不适用于存储大规模数据，这不仅会造成内容过度占用，而且在同步共识过程中也会浪费大量网络资源与计算资源。因此，按链上索引存储和链下数据存储 2 部分设计存储架构。链上存储用户信息、元数据信息及数据索引信息。区块由区块头和区块体构成，区块头里存储上一个区块的哈希值、本区块体的哈希值及时间戳等信息；区块体中存储用户、元数据及索引的加密信息，并通过区块间的连接进行区块的同步。对其他的业务数据，如客货运输部分数据，将其加密存储于 IPFS 中，称之为业务数据的链下存储。

（2）区块层。区块层为铁路数据汇集共享体系提供区块链核心技术功能，依托区块链技术提供多节点的身份认证和管理，在数据汇集共享应用场景中建立底层的交互协议。其中既包括各类共识算法、隐私权限策略、网络通讯策略，也包括数字签名、多重加密、非对称加密等贯穿铁路数据汇集共享中各环节的关键技术。

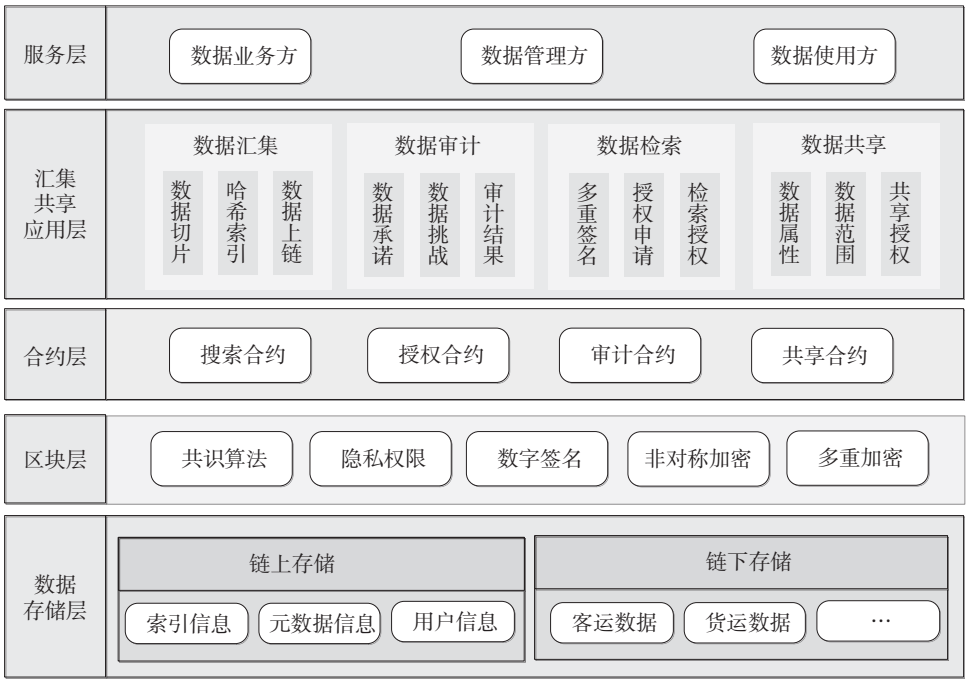

图 1　基于区块链技术的铁路数据汇集共享体系架构

Fig.1　Railway data collection and sharing architecture based on blockchain

（3）合约层。合约层集成铁路数据汇集共享的各类智能合约[7]，建立可监管、可审计的合约形式化规范。铁路数据汇集共享体系架构的合约层主要包括检索合约、授权合约、审计合约与共享合约。其中，审计合约在数据完整性审计过程中执行数据挑战确保数据的完整性；授权合约用来将数据部分特征如元数据信息、审计信息等授权给特定用户；检索合约在数据检索过程中判断用户是否具备检索权限，对于有权限的用户从区块链账本中检索；共享合约在数据共享过程中判断用户是否具备共享权限。

（4）汇集共享应用层。汇集共享应用层包括了铁路数据汇集共享中数据汇集、数据审计、数据检索和数据共享 4 项关键应用。①数据汇集。在数据汇集中，数据汇集方对汇集数据进行合理切片，将数据切片、元数据信息、数据摘要信息等存入 IPFS，并将 IPFS 返回的哈希索引存储在区块链中实现数据上链。②数据审计。在数据审计中将完成对汇集数据的完整性审计，首先数据汇集方根据数据特点提出数据承诺，数据审计方根据承诺构建挑战信息组，智能合约利用挑战信息组对被审计数据执行挑战，数据审计结果被存储于区块链中。③数据检索。数据检索可以在使用汇集数据前检索数据的粒度、属性和范围，分析数据是否满足业务需求。用户对数据审计信息和数据查询发起授权申请，授权通过后授权信息将被记入区块链全网广播，具体授权过程将通过授权合约完成，多重签名被引入授权以保证数据安全。④数据共享。在数据共享中，用户可以针对特定的属性、范围发起数据共享申请，数据汇集方和管理方进行申请审批并将授权信息写入共享合约，用户根据授权信息在 IPFS 获得共享数据，共享权限可以被随时收回以保证汇集方对业务数据的完全掌控。

（5）服务层。服务层中包括数据汇集共享体系中服务的 3 类用户，即数据业务方、数据管理方、数据使用方。①数据业务方。在铁路数据业务中，通常包含专业信息系统的建设运维单位及业务主管部门 2 类角色，前者从技术层面进行数据的运营维护，后者对数据的使用及共享进行审批管理。为了业务流程的简化，将这 2 类角色合并为数据业务方一个实体，数据业务方在本方案中既完成数据汇集，也作为数据管理者决定数据的查询共享范围。②数据管理方。数据管理方对汇集的数据进行整体管理，包括对数据业务方汇集的数据进行完整性审计，同时参与数据查询共享申请的审批。③数据使用方。数据使用者是需要数据的铁路其他业务部门，数据使用者在获得数据业务方和数据管理方的同意后，通过重建哈希的方式与区块链节点交互以证明其凭证，从 IPFS 节点获取数据。

2　基于区块链的铁路数据汇集共享关键技术

铁路数据汇集共享中根据业务流程需解决数据汇集、数据审计、数据检索、数据共享 4 个环节的技术问题。

2.1　铁路数据汇集技术

铁路数据汇集中的数据业务方首先对汇集数据进行合理切片。以货票为例，通常包含时间、发到站、运输经由、运价里程、货物名称等多个属性，既可以按属性将数据进行横向切片，也可以按不同的时间进行纵向切片，切片方式将根据数据特点及可能共享的方式决定，切片的粒度将影响后续数据共享的权限管控粒度。

铁路数据业务方将生成数据切片以及数据的元数据信息及数据摘要信息存入 IPFS 中并获取 IPFS 返回的哈希。其中数据切片用于数据共享，元数据信息用于数据检索，数据检索与共享的分离将提高数据安全性。哈希被数据业务方秘钥 SK^o 与数据管理方秘钥 SK^g 进行多重加密后交给其他区块链节点，区块链节点收到请求首先验证数据业务方与数据管理方的签名，通过后执行智能合约将哈希值广播到其他节点执行相同的操作，完成数据上链，数据管理方建立数据类别索引信息。基于区块链的铁路数据汇集如图 2 所示。

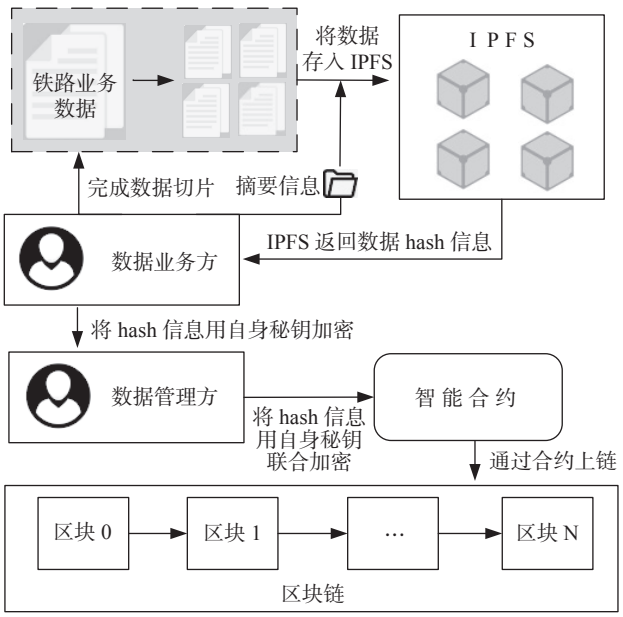

图 2　基于区块链的铁路数据汇集

Fig.2　Railway data collection based on blockchain

2.2　铁路数据审计技术

中国国家铁路集团有限公司 (以下简称“国铁集团”) 提出在数据汇集中要建立数据治理机制，制定数据标准规范，加强基础数据管理。因此，数据审计是铁路数据汇集共享中的重要环节，也是保证数据质量的关键。在数据业务方将业务数据存入 IPFS 后，数据管理方需要对数据的完整性进行审计，传统审计方法需要将全部数据暴露，需要找到一种既可以保护数据隐私又可以完成数据完整性审计的方法，引入零知识证明来解决这个问题。

铁路汇集数据的完整性审计过程如下：首先数据业务方根据数据特点提出一组包含 n 组承诺的承诺集 $T_c = (C_1, C_2, \cdots, C_n)$，并提出对应的承诺属性集 Z_d^*，将承诺集用数据管理方的公钥 PK^g 加密后写入区块链，数据管理方私钥 SK^g 解密该承诺集并审计认可后用私钥 SK^g 签名并写入数据审计智能合约。在数据审计时，数据管理方随机选择 m ($m<n$) 个随机数 $i\in[1, n]$ 和对应的承诺属性 $v_i\in Z_d^*$，基于 (i, v_i) 构造挑战信息组，将挑战信息组发给智能合约，智能合约对被审计数据执行挑战，得到被审计数据集的持有性证明集 $P_s = (H_1, H_2, \cdots, H_m)$，式中 H 为针对各属性的持有性证明。通过校验 P_s 与 T_c 完成数据审计，证明数据业务方已完成数据汇集。铁路数据汇集中的数据审计结果被数据管理方通过私钥 SK^g 签名后写入区块链中。基于区块链的铁路数据审计总体流程如图 3 所示。

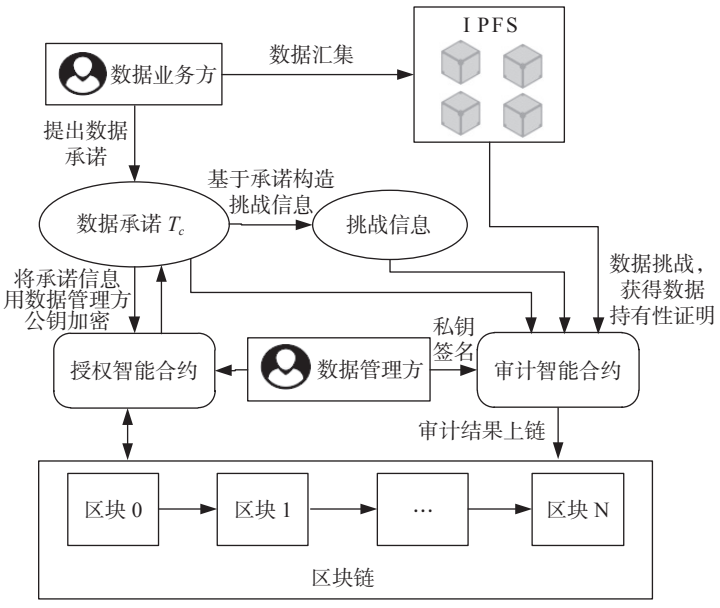

图 3　基于区块链的铁路数据审计总体流程

Fig.3　Overall process of railway data audit based on blockchain

2.3　铁路数据检索技术

铁路数据中，业务部门在使用其他汇集数据前

首先需要检索数据信息。例如，机务系统中某个部门计划开展全路机车故障的数据分析，需要使用汇集数据中的机车电子档案信息，在申请共享前应当先了解机车电子档案中包含哪些信息及具体信息的粒度、属性和范围，分析其是否满足机车故障分析的业务需求。

将数据检索与数据共享分离开，目的在于规范铁路数据共享流程、保障数据安全。数据检索被设计成需要数据业务方和数据管理方双方许可才能被执行，这个过程中使用了区块链多重签名技术。铁路中一个业务部门可能管理着某一领域多类数据，为了保证数据安全性，使用中一次一密的椭圆曲线加密方案[8]，数据业务方将自己的私钥 SK^{ot} 通过 $PK^{ot} = SK^{ot}\ G$ 计算得到针对某类数据 t 的公钥 PK^{ot}，式中 G 为椭圆曲线的一个基点。之后利用数据管理方的公钥 PK^{g} 和 PK^{ot} 共同得到数据 t 的联合签名地址 Dt，$D_t = H(PK^{g},\ PK^{ot})$，其中 H 为哈希函数。

数据使用方对具体铁路数据 t 的检索流程如下：首先数据使用方申请查看数据 t 审计情况，数据管理方许可后将区块链上的数据 t 审计信息授权给数据使用方，签名后发起数据授权请求。区块链节点验证该请求，执行智能合约后广播该请求，最后将区块链上的数据 t 审计信息用数据使用方的公钥加密后发送给数据使用方，数据使用方用自己秘钥解密后即可查看审计信息。确认数据审计结果后，数据使用方通过联合签名地址 D_t 发起数据查询授权申请，针对这一申请，数据业务方和数据管理方分别用自己的秘钥 SK^{ot} 和 SK^{g} 完成多重签名授权后，区块链节点进行该授权的有效性检查，通过后记入区块链并进行广播。

数据使用方在发出对数据 t 的查询请求后，智能合约将根据授权信息和查询数据类别判断用户是否具备权限，对于有权限的用户从区块链账本中检索出数据 t 的元数据信息及数据摘要信息对应的哈希记录，数据使用方根据哈希记录向 IPFS 节点发起请求完成数据查询。

2.4　铁路数据共享技术

在铁路数据共享中，如何实现对数据共享的精细管控是重点。精细管控包括 2 个方面的内容，首先是精细化管控共享数据的属性和范围，如在货运客户分级评价和流失预警中需要获取过去特定时间段中货主的发货量、发货品类等属性信息，超出范围的货主信息如个人身份信息等并不应该被共享。其次是在这个过程中，数据业务方可以随时收回业务数据共享权限，保证对数据的完全掌控。

数据使用方完成业务数据检索后，可以针对数据特定的属性或范围发起共享申请，与数据检索相似，数据使用方通过联合签名地址 D_t 发起数据共享授权申请，数据业务方进行审核后用自己的秘钥 SK^{ot} 签名授权，数据管理方同样需对该申请进行签名授权，区块链节点验证该双重签名后将授权信息，包括授权的数据业务方 ID、数据使用方 ID、授权的数据属性与范围写入数据共享合约。区块链节点调用数据共享合约，如果不满足授权条件将忽略该次请求，如果满足授权条件，将授权属性及范围内的数据在 IPFS 的哈希信息用数据使用方的公钥加密后发给使用方，数据使用方使用自己的秘钥解密后在 IPFS 获得共享数据。

2.5　铁路数据汇集共享体系架构安全特性分析

（1）存储安全。首先，针对铁路数据特点采用数据链上链下存储分离的方式，链上存储元数据信息及数据索引信息，数据本体被加密后存储于链下分布式文件系统中，链上与链下节点的分布特点确保了部分节点失效不会造成整个网络的瘫痪。其次，各专业数据索引信息被多重加密后存储于区块链上，数据共享通过智能合约进行交互，未经允许的获得索引数据无法查看其中的信息，这些做法提高了整体安全性，降低了数据泄露的风险。

（2）防篡改。采用 IPFS 存储与区块链结合的方式，所有的节点都保存完整的区块链数据，在 PBFT 共识机制下，只要保存链上数据节点被篡改的数量不超过总节点数的 1/3，经过一轮共识后正确的数据将得到恢复；同时线下数据存储在 IPFS 中，对象之间的链接关系通过 hash 加密被保存，当链下数据被篡改后也将被立刻校验检测出来，确保存储的汇集数据难以被篡改。

（3）可追溯。通过数字签名的形式记录了铁路数据共享过程中汇集、审计、授权、共享等过程的全部信息，可以在数据视角上实现铁路数据汇集中

“谁汇集、谁授权、谁申请、谁使用”的全流程溯源，规范了铁路数据汇集共享流程，形成铁路数据资产汇集应用的全景视图。

（4）权限管理。数据权限的管理不简单的依赖于平台管理者，通过多重授权机制，各类业务数据的管理部门可以控制自己数据的查询和共享权限，有效掌握业务数据的使用情况。同时依托 IPFS 分布存储中对数据的横纵切分，实现对数据共享的精细管控。

3　结束语

铁路数据汇集共享中，既要汇集整合全路数据资源，实现数据的规范存储、管理和高效应用，又要解决数据业务方对敏感数据的共享管控。基于区块链技术的铁路数据共享架构将数据共享与存储分离，既解决了大数据上链带来的低吞吐高延迟问题，又从数据汇集、数据审计、权限管理、数据共享等多个方面提出技术解决途径。基于区块链技术的铁路数据共享架构不但可以用于铁路内部，还可以推广应用到海铁联运、客运延伸服务等铁路与其他行业的数据共享交换之中。未来可以引入 token 即代币，通过代币机制激励数据流通。

参考文献：

[1] 袁　勇，王飞跃. 区块链技术发展现状与展望[J]. 自动化学报，2016，42(4)：481−494.
YUAN Yong，WANG Feiyue. Blockchain：The State of the Art and Future Trends[J]. Acta Automatica Sinica，2016，42(4)：481−494.

[2] 王洪业. 基于区块链技术的空铁联运售票模式研究[J]. 铁路运输与经济，2019，38(11)：121−124.
WANG Hongye. A Study on Air-Rail Inter-modal Transport based on Block Chain Technology[J]. Railway Transport and Economy，2019，38(11)：121−124.

[3] 杨　明，丁　龙，许　艳. 基于区块链的医疗数据云存储共享方案[J]. 南京信息工程大学学报 (自然科学版)，2019，11(5)：590−595.
YANG Ming，DING Long，XU Yan. A Cloud Storage and Sharing Scheme for Medical Data based on Blockchain[J]. Journal of Nanjing University of Information Science and Technology (Natural Science Edition)，2019，11(5)：590−595.

[4] GAI K，CHOO K R，WU Y L，et al. Controllable and Trustworthy Blockchain-based Cloud Data Management[J]. Future Generation Computer Systems，2019(91)：527−535.

[5] 张利华，蒋腾飞，姜攀攀，等. 基于区块链的高速铁路监测数据安全存储方案[J]. 计算机工程与设计，2020，41(4)：933−938.
ZHANG Lihua，JIANG Tengfei，JIANG Panpan，et al. Secure Storage Scheme for High-Speed Railway Monitoring Data based on Blockchain[J]. Computer Engineering and Design，2020，41(4)：933−938.

[6] XIA Q，SIFAH，EMMANUEL B，et al. MeDShare：Trust-less Medical Data Sharing among Cloud Service Providers via Blockchain[J]. IEEE Access，2017(5)：14757−14767.

[7] DELMOLINO K，ARNETT M，KOSBA A，et al. Step by Step towards Creating a Safe Smart Contract：Lessons and Insights from a Cryptocurrency Lab[C]// Proceedings of the 2016 International Conference on Financial Cryptography and Data Security，LNCS 9604. Berlin：Springer，2016：79−94.

[8] DAI M，ZHANG S，WANG H，et al. A Low Storage Room Requirement Framework for Distributed Ledger in Blockchain[J]. IEEE Access，2018(6)：22970−22975.

收稿日期：2020−07−08
基金项目：中国国家铁路集团有限公司科技研究开发计划课题 (K2019S012)
责任编辑：张婷钰