计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210100163

基于区块链的大数据交易关键技术与发展趋势

曹 萌[1,2] 于 洋[1,2] 梁 英[1] 史红周[1]
1 中国科学院计算技术研究所 北京 100190
2 中国科学院大学计算机科学与技术学院 北京 100190
(caomeng19s@ict.ac.cn)

摘 要 大数据时代下各类数据价值日益凸显,不同主体对大数据交易的需求也愈加迫切。传统集中式平台下的大数据交易存在用户数据被恶意采集、隐私泄露、数据被转售、数据虚假等诸多风险。一般认为,采用具有去中心化、透明性、隐私保护、不可篡改性等特点的区块链技术是解决上述大数据交易问题的一种重要途径。然而,区块链技术在大数据交易领域的应用还处于早期发展阶段,应用方案尚未成熟。对此,对目前学术界提出的多种基于区块链技术的数据交易方案进行总结,从隐私保护、数据转卖和交易公平3个角度出发,介绍使用区块链技术提升集中式数据交易的具体方法,并对各方法的优缺点进行分析;最后从隐私保护、身份认证、海量数据等方面分析"区块链+大数据交易"目前面临的挑战和未来的发展方向。

关键词:区块链;大数据交易;隐私保护;访问控制;身份认证
中图法分类号 TP311

Key Technologies and Development Trends of Big Data Trade Based on Blockchain

CAO Meng[1,2], YU Yang[1,2], LIANG Ying[1] and SHI Hong-zhou[1]
1 Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
2 School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100190, China

Abstract In the era of big data, the value of data has become increasingly prominent, and the needs of big data trading for different entities are becoming more and more urgent. Traditional big data trading based on centralized platforms have many risks such as malicious collection of user data, privacy leakage, data resale, and data fraud. It is generally believed that using blockchain technology with the characteristics of decentralization, transparency, privacy protection, and tamper resistance is an important way to solve the above problems in big data trading. However, the development of blockchain technology in big data trading is still in the early stage, and the application plans are not yet mature. In this regard, this paper summarizes the current data trading models based on blockchain technology proposed by the academic community, introduces the methods of improving centralized big data trading from the three perspectives of privacy protection, data resale and trading fairness by using blockchain technology, and analyzes the advantages and disadvantages of each method. Then, this paper analyzes the current challenges and future development directions of "blockchain+big data trading" from the aspects of privacy protection, identity authentication, massive data, and value transfer.

Keywords Blockchain, Data trading, Privacy protect, Access control, Identity authentication

1 引言

随着大数据时代的到来,世界数据总量呈现爆炸式增长,尤其是物联网及各类智能物件[1-3]逐渐融入人们的日常生活,数据规模还会进一步扩大[4]。《2020中国大数据产业发展白皮书》指出,2019年中国大数据产业规模达5397亿元,预计2022年整体规模突破万亿元。随着数据挖掘和机器学习技术[5]日渐成熟,数据价值日益凸显,数据交易逐渐兴起,数据交易的发展一方面使得数据所有者获得更具个性化的服务,另一方面又使数据购买者提高收益[6]。

而数据作为一种数字资源,明显不同于传统商品,数据具有复制成本低、增长速度快、价值难以估量等特点,这些特点使得数据在交易过程中面临许多挑战。在集中式数据交易平台下,用户一旦将数据上传就失去了对数据的控制权,存在被平台非法使用和出售的风险;即使平台完全可信,也会因为平台提供方的单一性存在单点故障,系统被攻击导致隐私泄露等风险。如2020年Facebook因未经用户同意利用其照片标记功能来收集用户的生物特征图像,导致用户隐私泄露,被罚款6.5亿美元。

区块链技术具有去中心化、不可篡改、可追溯等特性,为解决数据交易中的上述问题提供了新思路:利用区块链建立一个无第三方的数据交易平台,在交易双方建立信任机制,交易过程可以采用加密技术保护用户隐私,使用区块链存储交易信息保证数据所有者对数据始终具有控制权,跟踪数据使

基金项目:国家重点研发计划(2018YFB1004700)
This work was supported by the National Key Research and Development Progra of China(2018YFB1004700).
通信作者:史红周(hzshi@ict.ac.cn)

用情况。基于这些优势，基于区块链技术的数据交易市场在业界掀起了一股研究和应用热潮，如国外的 IOTA IoT 数据市场[7]、DatabrokerDao[8]和 BAIC[9]等，国内的 GXB[10]、上海数据交易中心、贵阳大数据交易所等。

然而区块链技术在大数据交易领域的应用还处于早期发展阶段，缺乏相关的研究文献，大部分关于数据交易的研究和应用分散在区块链在物联网[11-13]、医疗[14-15]、能源、金融[16]等领域应用文献中。本文针对这一现状，对基于区块链技术的数据交易模式进行综述分析，并讨论区块链和数据交易结合所面临的挑战和可能的解决方法，希望给基于区块链的数据交易发展提供一定的理论支持。

本文第 2 节简要介绍集中式数据交易的发展现状、区块链技术和区块链特性在数据交易中的应用；第 3 节从数据交易面临的 3 个问题出发，对比集中式交易模式和区块链下交易模式的解决方法，分析其优缺点；第 4 节重点探讨“区块链＋大数据”面临的挑战和未来的发展方向；最后总结全文。

2　背景

2.1　集中式数据交易研究现状

当前集中式交易模式主要有两种：托管模式和聚合模式[17]。托管模式是数据所有者将自己的数据托管到交易平台，采用这种模式，交易平台将获得数据的控制权，数据所有者将对数据后续的使用和交易失去控制，可能存在数据所有者利益受损的问题。采用聚合模式可以一定程度缓解此类问题，在聚合模式中数据交易平台通过 API 接口连接数据，数据仍然由数据所有者自己管理。但采用聚合模式在数据流经平台时，平台仍有机会留存数据，且存在单点故障问题。

无论托管模式还是聚合模式的数据交易机制，都正随着不断的应用实践而成熟，但集中式的数据交易模式始终面临着各种问题。

如图 1 所示，在集中式交易模式下，数据卖方和买方通过数据交易平台进行交易，数据卖方发布信息时的身份校验由交易平台负责，发布信息也由第三方平台存储，这样在技术上为平台篡改或盗取数据卖方的数据信息提供了可能；且现有的数据交易平台也很少为数据所有者提供数据跟踪功能，数据所有者对数据没有知情权和控制权；此外，平台无法判断用户发布的数据是否为非法获取或是意图转卖数据。

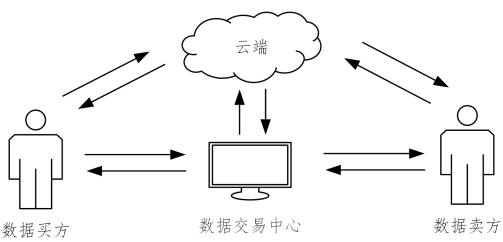


图 1　集中式数据平台架构图[18]

Fig. 1　Centralized data trading platform diagram[18]

2.2　区块链

区块链是从比特币底层技术衍生出来的新型技术，最早的定义来自于中本聪 2009 年发表的论文[19]，它是一个由多方共同参与维护的分布式数据库，每一个参与维护的节点都保存了完整的区块链数据，这些节点之间通过运行一致性算法，如工作量证明 POW、股份证明 PoS 等形成共识，实现对账本状态的改变。区块链的存储结构如图 2 所示，每个区块分为区块头和区块体，区块头存储上一区块的哈希值，使得已经写入区块的数据很难被篡改，区块体存储详细的交易数据。

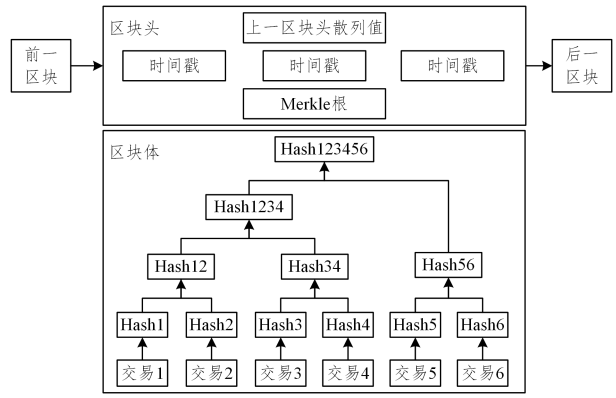


图 2　区块链数据结构图

Fig. 2　Blockchain structure diagram

中本聪的论文为区块链提供了思想，所以被称为区块链 1.0，之后逐渐出现区块链 2.0 以太坊[20]，数字货币与智能合约的结合。智能合约是指运行以太坊虚拟机上（EVM）上的一段代码，它将比特币只能实现比特币交易的功能进行拓展，可以在无第三方的情况下进行安全的交易。

近年来，区块链概念逐渐被应用于供应链[21]、金融交易[22-23]和物联网[24]等各个领域。

2.3　区块链与数据交易

区块链的多节点性、不可篡改性、强制执行性为解决数据交易中存在的问题提供了思路。

1）多节点性：在区块链中进行的身份认证、数据交易和数据验证等过程都经过多个节点的验证，可以解决第三方服务平台的不可信问题。

2）不可篡改性：使用区块链存储数据的访问控制列表和数据交易记录，可以提高访问控制的安全性，并为数据所有者提供真实可信的交易日志。

3）强制执行性：数据代理和数据请求者使用智能合约制定交易规则，智能合约生效后自动执行，不被任何一方干预，为买卖双方建立信任关系提供了技术基础。

3　区块链下的数据交易

本部分将从隐私保护、数据转卖和交易公平 3 个方面详细介绍如何利用区块链对集中式数据交易模式进行提升。

首先定义参与交易的多个实体（这些实体都通过生成公私钥对的方式向区块链注册）。

1）用户（数据所有者）：产生数据的实体。

2）数据代理：一般数据所有者拥有少量数据，可以通过数据代理整合处理数据后向数据请求者出售，数据代理也可以是数据所有者本身。

3）数据请求者：查询或购买数据的实体。

3.1　隐私保护

数据交易中的隐私泄露体现在两个方面，一方面是指数据代理未经用户同意就收集用户数据或是将数据出售至用户未许可的数据请求者，另一方面是指数据代理未对数据中的隐私信息进行处理和保护，导致用户隐私信息泄露。

在集中式的数据交易模式下，大多数据代理混淆数据所

有权和数据使用权，以服务换数据，未经数据所有者同意就采集和出售数据，导致数据所有者失去对数据的知情权和控制权。对此，只能通过法律手段对数据代理和交易平台进行约束，但法律手段只能用于用户发现隐私泄露后的维权，并不能从技术上避免这种现象的发生。

在基于区块链的数据交易模式下，数据代理发布数据时，区块链节点需要验证该数据已通过数据所有者授权才允许写入区块链。数据交易后，需要将交易记录存储在区块链上，确保用户的知情权和控制权。此外，区块链的匿名性也可以保护数据交易双方的身份信息。

Kiyomoto 等[25]提出使用隐私管理器(PPM)实现这一机制。用户通过 PPM 设置自己的隐私偏好，当数据代理的请求行为违背该偏好时，PPM 将会向用户发送请求信息，用户来决定是否允许该请求。若用户同意，则将数据发送给数据代理，数据代理对数据进行处理，生成符合交易平台出售规则(如符合 k 匿名)的匿名数据集 D，并使用其 hash 值作为数据集 ID 发送至 PPM。此后，数据代理和数据请求者关于该数据集的数据交易都会被区块链验证和存储，用户可以通过数据集 ID 查看相关交易日志。由于区块链的公开透明和不可篡改等特性，可以为用户提供真实可靠的数据跟踪功能，该过程如图 3 所示。

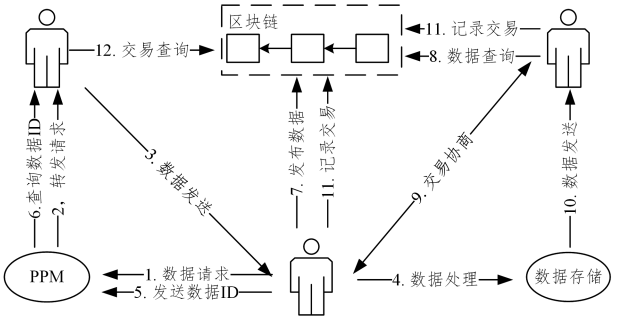


图 3 数据交易流程图

Fig. 3 Data trading process diagram

该机制解决了隐私泄露中的第一个问题，使用户对数据具有控制权和知情权，但是未解决第二个问题，无法确保数据代理在收到数据时按照平台规定生成相应的匿名数据集，对此平台无具体的检测机制。

Zhou 等[26]提出，由用户直接出售数据，用户将数据使用 AES 加密上传到服务器，并将数据索引放入区块链，数据请求者检索并购买数据，整个过程不需要第三方参与，实现自动买卖。但仅通过数据索引，不能使请求者判断数据是否满足要求，数据请求者在购买之前需要了解特定的数据信息，如“这个病人在过去 3 年，有过多少次肾衰竭的经历”，再决定是否购买。区块链平台需要访问整个待交易数据，这将带来数据泄露风险。为此，文章提出使用数据嵌入，将原始数据映射到低维的隐私保护签名中，使用高性能的投影函数，确保投影后的向量可以较全面地描述原数据。查询数据时，使用与特定任务紧密相关的相似性函数[27]，将签名向量作为函数输入来判定检索出的数据条目的相似性。在该方法中，用户可以自由出售数据内容，实现对数据的控制权。但数据出售以用户为单位，对于大数据分析来说数据量较少，需要多次购买，过程较为繁琐，数据嵌入和数据查询对计算能力要求较高。

3.2 数据转卖

数据由于其特殊性——“所见即所得”，无复制成本，数据转卖成为数据交易中的一大难题。数据存储方、交易平台和数据请求者在获得数据后都可能对数据进行转卖，造成数据所有者利益受损。

在集中式的数据交易模式下，一些数据所有者选择对数据加密，这一方法虽然可以避免数据存储方和交易平台对数据进行转卖，但仍然无法规避被数据请求者再次转卖的风险。为此，一些交易平台[28]设置数据检测机制，将数据代理提供的数据与之前存在的数据进行对比检测，若重合率较高，则判定为数据转卖，但购买者只要对数据进行任意修改就可以绕过检测。

在基于区块链的数据交易下，可以要求数据代理向区块链注册数据时提供数据来源，采用区块链技术可以对数据进行溯源，无论经过多少次复制、传播，都可追溯数据生产者和所有者，但该机制并不能阻止数据请求者在得到数据后在其他平台转卖数据。

Dai 等[29]提出一种基于区块链的安全交易方法 SDTE，交易中数据代理出售数据分析结果而非原始数据。首先，数据请求者以智能合约的形式发布数据处理算法至区块链，广播购买需求，并从响应节点中选择合适的数据代理和可信节点作为数据源和数据处理节点。可信节点指受 SGX[30]保护的 EVM 节点。SGX 是英特尔架构的一种扩展，允许在内存中创建一块区域，应用程序向其中传入代码和数据，只有 CPU 和该应用程序可以访问，在硬件层次上确保数据安全。选定节点后，数据请求者将可信节点信息发送给数据代理，数据代理将数据地址和解密密钥发送给可信节点。可信节点执行处理过程，将结果使用数据请求者提供的加密密钥加密，并发送给数据请求者。

该方法中数据请求者只得到数据处理结果，从根本上解决了数据转卖问题，同时为数据请求者解决了计算问题。数据处理过程在可信环境中执行，确保分析过程中原始数据不泄露，保证数据安全。但该机制未对智能合约进行限制，数据请求者可能编写特殊的智能合约使其直接输出原始数据，对此文章提出使用计算费用与输出结果大小相关的规定，但数据请求者和可信节点联合就可避免高额费用。

3.3 交易公平

交易公平是数据交易的最后一道保障，要求数据交付和费用支付是“原子性”的，确保交易最终结果要么是数据请求者收到合法数据，数据代理收到付款；要么是数据请求者未收到或收到无效数据，数据请求者无需支付。

在集中式交易平台中，当数据有效性出现争议时，往往需要第三方平台介入，难免会出现主观上的偏袒，且由于数据的特殊性，一旦查看就很难“退货”。

在基于区块链的数据交易模式下，数据的有效性和完整性使用多节点检验，一定程度上提高了公正性。数据交易的“原子性”可以由智能合约来保障。

为实现这一机制，Xiong 等[31]提出在交易前，数据代理和数据请求方需要在智能合约中支付与数据等金额的保证金，交易成功后智能合约将返还数据代理的保证金和规定比例的数据金额。如果交易存在争议，多个节点会对数据有效性进行验证，计算数据代理描述的特征向量与请求者收到数据的特征向量的马氏距离(Mahalanobis Distance)[32]，如果验证未

通过，智能合约将不会返还数据代理的保证金。该方案可以确保数据的有效性和交易的完整性。

Zhao 等[33]提出使用双重签名技术(DAPS)[34]来保证交易的完整性，交易时数据代理向数据请求者发送第一个 DAPS 签名，请求者向区块链发布支付合约，数据代理调用支付合约将 DPAS 的第二个签名发送到区块链以接收付款，数据请求者通过两个 DAPS 签名提取密钥，解密数据。如果数据请求者未完成支付，则无法获得第二个签名；区块链各个节点只能得到第二个签名，无法获取第一个签名，无法提取密钥。该方案使用较为简洁的方式在保护数据安全的同时，保证了数据交易的完整性。

综上所述，我们对于数据交易中存在的隐私保护、数据转卖、交易公平等问题，将集中式平台下的解决方案与区块链平台下的解决方案进行总结对比，如表 1 所列。

表 1　数据交易问题解决方案的对比

Table 1　Comparison of solutions to data trading problems

	数据交易常见问题	集中式交易平台现有解决方案	基于区块链的交易平台解决方案
隐私保护	数据代理未经同意采集用户数据；数据实际出售范围超出用户规定	技术上未实现，由法律约束	在区块链发布数据时需要多节点验证用户的授权证书；用户通过区块链跟踪数据交易记录
	数据代理出售含用户隐私的数据	技术上未实现，由法律约束	用户直接出售数据
数据转卖	数据存储方转卖数据	将数据加密后存储	将数据加密后存储
	数据请求者转卖数据	数据代理仅出售数据分析结果	数据代理仅出售数据分析结果；出售数据时区块链需要验证数据来源
交易公平	数据交付与支付的完整性	交易平台监督	使用智能合约自动执行交易；双重加密技术
	交付数据的完整性和有效性	交易平台验证	使用区块链中关于数据的 hash 值等元信息验证；在区块链多个节点使用距离度量算法比较相似性验证数据有效性

通过对比发现，在集中式交易平台下，问题解决大多依靠第三方平台，无法在各参与实体间建立较强的信任关系。而在基于区块链的交易平台下，可以利用区块链的多节点性、透明性和不可篡改性，结合密码学、机器学习等技术解决问题，增强用户对平台的信任。

4　问题与发展方向

将大数据交易与区块链技术结合，可以为数据交易现存问题的解决提供新思路，但区块链自身也存在一些弊端，如可扩展性差、能耗高[35]、易遭受女巫攻击[36]等。研究人员通过各种方式不断克服以上问题，如使用链上扩容(扩块、分片、改进共识机制等)[37]和链下扩容(状态通道、侧链等)来增强区块链的可扩展性；改进共识算法来降低达成共识消耗的能耗；使用 P2P 节能协议[38]或微型链[39]减少通信能耗；对向区块链发布请求的实体收取费用或是采用许可网络 fabric 来抵御女巫攻击等。

除此之外，本部分将着重分析“区块链＋大数据交易”所面临的挑战，包括隐私保护、海量数据、身份认证，并给出可能的优化建议；最后提出未来发展方向。

4.1　隐私保护

区块链在隐私保护方面的弊端和前文提到的优势并不冲突。优势是指区块链与集中式平台相比，可以保证数据所有者对数据的知情权和控制权，不存在第三方隐藏交易或篡改数据；其弊端是指区块链存储的交易信息和访问控制列表是公开透明的，各个节点都可访问(因为区块链的信息要被所有节点验证)，各实体尽管大多以匿名(实际上是假名)的方式加入区块链，但被证明仍有被获取真实信息、追踪到真实个体的可能[40]。

解决区块链带来的隐私问题主要有两种方式[41]。第一种是在现有区块链中增加匿名化，常用的有采用 CryptoNote[42]环签名技术。环签名技术中，一组用户(环)生成签名，验证者只能验证签名是属于这一组用户中的一位，但并不能确定是哪一位，以此隐藏用户的身份信息，使交易无法通过区块链跟踪；采用同态加密[43]、多方安全计算技术使各节点完成验证的同时无法得知验证的内容；或是实体每参与一次交易就更换一次公私钥对。使用这些技术应考虑计算资源，尤其是在计算资源受限的设备上。第二种方式是创建对现有区块改进的区块，如 zero-cash[44]使用零知识证明 ZK-SNARK 技术，可以实现交易双方、交易内容的完全保密；fabric 使用多通道技术，通道之间完全隔离，但通道内部成员仍然可以看到交易内容，fabric 的隐私保护要保证共识节点可信；也有人提出使用混合方案增加隐私保护，将来自不同用户的交易混合在一起，达到匿名效果，但需要依靠第三方混合，这样会引入不信任实体。

4.2　身份认证

在数据交易中，各方实体向区块链发布事务时，数据访问控制都会用到身份认证技术，身份认证技术也在很大程度上影响着平台的安全性和可靠性。

在集中式数据交易模式下，身份认证由第三方负责，认证过程较为简单，用户输入正确的用户名密码即可通过验证。在区块链中，身份认证需要在多个节点进行，使用比较广泛的有传统公钥基础(PKI)体系，通过证书验证身份，需要管理大量证书，尤其是区块链需要多节点验证，会产生信息冗余；基于身份的密码体制(IBC)，可以通过用户公钥验证身份，但用户私钥由密钥生成中心(KGC)生成，会产生密钥托管问题。

LI 等[45]提出使用无证书密码技术[46]，用户私钥由用户的身份和 KGC 不知道的某个参数生成，在解决密钥托管问题的同时避免了传统 PKI 技术带来的冗余；此外，区块链系统的公共账本为广播公钥提供了一种便捷方式，克服了无证书密码体制的缺点。

但在这种管理访问控制时，会出现节点资源受限问题。OUADDAH 等[47]提出了一种去中心化的匿名和隐私保护授权管理框架 FairAccess，在区块链中以事务形式存储每对(资源、请求者)访问控制策略，引入比特币中 Wallet 的概念，为不同设备安装自己的 Wallet。Wallet 起到访问控制代理的功能，资源拥有者创建事务将访问控制策略加载到区块链，同时

向被授权的访问请求方账户发送授权令牌的形式进行权限管理，令牌由资源拥有者使用其私钥进行签名来保证其不可伪造，通过这种双重开销检测机制检测令牌重用。

4.3 海量数据

大数据时代下数据量不断增加，数据加密存储和流通难度变大，尤其是在区块链技术下弊端更为明显，若将数据存储在区块链中，存储代价较高，可扩展性较差，而且一旦加密算法被攻破，区块链的透明性和不可篡改性将导致之前存入区块链的数据被全部泄露且无法挽回[36]。若将数据存储在第三方，存在数据被修改的风险，且意味着访问控制需要由第三方实现。

Liang 等[48]提出在云端存储时使用 Merkle 树[49]结构存储数据。Merkle 树是一种二叉树，结构与图 2 区块链中的区块体结构相同，每两个记录为一组，两个记录的哈希值串联生成其父节点，一直重复直至生成根节点，并将哈希值存入区块链。当其中一个节点被篡改，遍历树的各个节点就可检测出被篡改的记录，这样实现了第三方存储的完整性验证，但是节点数据未被加密，存在被数据存储服务方窃取的风险。

Li 等[45]提出利用边缘计算将用户在物联网设备中的数据加密存储在分布式哈希表(DHT)[50]中，并将数据地址写入区块链。当数据请求者 A 想要访问数据时，需要数据所有者将 A 的 ID 写入区块链事务中，DHT 节点查询区块链数据，若事务被记录，则将数据进行重加密发送给 A，确保用户 A 可以使用自己的私钥解密数据。重加密技术密[51]可以实现数据在一个密钥加密下直接转换为在另一个密钥下加密，无需对数据解密。这种方法避免了数据存储方窃取原始数据的风险，且实现数据安全交付，但是加密发送过程较为繁琐，且加密次数随着交易量而急剧增加。

Xiong 等[31]使用 IPFS 文件存储加密后的待交易数据。IPFS[52]是一种开源的、分布式、点对点的，存储大规模数据的有效方法，解决了数据存储的可扩展性。当数据请求者购买成功后获得令牌，数据存储节点通过验证令牌的有效性准许请求者下载数据。这种方法降低了计算复杂性，但数据存储和访问控制由第三方负责，会降低系统的去中心化性和安全性。

通过上述几种方案，我们发现若将数据存储在区块链，由于多节点都要存储，会带来巨大的存储代价；若将数据存储在由第三方提供的服务中，就需要对数据加密，将解密密钥发送给数据请求者，但每次交易后需要更改加密密钥，造成较大计算量。因此，我们需要根据具体需求在存储与计算以及中心化和可扩展性之间取得平衡。

4.4 未来方向

一般基于区块链的数据交易平台，主要解决未经授权就采集和转售数据等问题。实际上，基于区块链的数据交易平台还可以支持数据所有者获得持续的数据交易利益，激励其更加主动地共享数据。

若在第三方平台实现该功能，不能保障数据代理诚实地根据销售情况向数据所有者返利。使用区块链平台，数据使用记录无法篡改，并可以利用智能合约实现自动返利。

使用区块链实现价值转移，不仅可以在数据所有者和数据代理之间建立信任，创造良好的数据共享环境，打破各行业间的数据孤岛，还可以利用区块链技术直接进行点对点交易的特点，实现点对点的价值直接转移，即在集成数据交易过程中将其中个体数据的价值直接转移至个体数据所有者，以此鼓励数据共享和流通。

例如，患者电子病历(EMR)已成为医疗大数据的重要内容，在医学研究、药物研发、医疗保险等领域具有重要价值，但电子病历数据本身不仅涉及医疗机构和医生，还涉及患者个人，其所有权、隐私保护、数据价值分配等在传统数据交易中很难做到各方兼顾。采用区块链技术的隐私保护以及价值转移机制，就可能很好地处理患者隐私、数据所有权、数据价值合理分配等方面的问题，可以实现患者在获得一定收益的情况下让渡部分隐私信息，推进医疗数据共享共用。

此外，为降低用户隐私泄露的可能，未来应更加鼓励对数据分析产品的交易，以降低大数据的使用门槛。此外，随着分析结果的多次出售，还可以缓解单次数据分析的计算成本。

结束语 现有集中式大数据交易面临诸多挑战，包括数据隐私保护、数据转卖、交易公平等，基于区块链的数据交易为大数据交易提供了一种新的交易模式，可以增强数据交易的安全性、隐私性、可追溯性和可靠性。本文研究了区块链与大数据交易的结合，分析了如何使用区块链解决大数据交易存在的上述问题，讨论了该模式中隐私保护、海量数据等挑战及可能的解决方法。

采用区块链的大数据交易技术在未来研究中，需要结合具体数据交易应用场景，充分发挥区块链技术特点，设计能更好体现数据价值保护、数据隐私保护、数据价值可控转移、数据分析产品交易、更易于落地实施的新机制和新方法，推动各类大数据的应用深度。

参 考 文 献

[1] SUN Y, SONG H, JARA A J, et al. Internet of things and big data analytics for smart and connected communities[J]. IEEE Access, 2016(4): 766-773.

[2] STANKOVIC J A. Research directions for the internet of things [J]. IEEE Internet of Things Journal, 2014, 1(1): 3-9.

[3] WU J, ZHAO W. Design and realization of winternet: From net of things to internet of things[J]. ACM Transactions on Cyber-Physical Systems, 2016, 1(1): 1-12.

[4] LIN J, YU W, ZHANG N, et al. A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications[J]. IEEE Internet of Things Journal, 2017, 4(5): 1125-1142.

[5] JAN B, FARMAN H, KHAN M, et al. Deep learning in big data analytics: a comparative study[J]. Computers & Electrical Engineering, 2019, 75: 275-287.

[6] GROVES P, KAYYALI B, KNOTT D, et al. The "big data" revolution in healthcare: Accelerating value and innovation[M]. Centre for US Health System Reform; Business Technology Office, 2013.

[7] IOTA blockchain data trading market[EB/OL]. https://data.iota.org/.

[8] Databroker Dao[EB/OL]. https://databrokerdao.com/.

[9] BAIC [EB/OL]. http://baic.io/.

[10] Gxb Blockchain White Paper [EB/OL]. https://cryptorating.eu/whitepapers/GXChain/GXB_Blockchain_White_Paper_v1.2_EN.pdf.

[11] LE NGUYEN B, LYDIA E L, ELHOSENY M, et al. Privacy preserving blockchain technique to achieve secure and reliable sharing of IoT data[J]. Computers, Materials & Continua, 2020, 65(1): 87-107.

[12] HUH S, CHO S, KIM S. Managing IoT devices using blockchain platform[C]// 2017 19th International Conference on Advanced Communication Technology (ICACT). IEEE, 2017: 464-467.

[13] REYNA A, MARTÍN C, CHEN J, et al. On blockchain and its integration with IoT. Challenges and opportunities[J]. Future Generation Computer Systems, 2018, 88: 173-190.

[14] XIA Q, SIFAH E B, SMAHI A, et al. BBDS: Blockchain-based data sharing for electronic medical records in cloud environments [J]. Information, 2017, 8(2): 44.

[15] CHEN Y, DING S, XU Z, et al. Blockchain-based medical records secure storage and medical service framework[J]. Journal of Medical Systems, 2019, 43(1): 1-9.

[16] CHEN Y, BELLAVITIS C. Blockchain disruption and decentralized finance: The rise of decentralized business models[J]. Journal of Business Venturing Insights, 2020, 13: e00151.

[17] YANG M J. Design of a data trading platform based on cryptography and blockchain technology[J]. Chinese Journal of Information Communication Technology, 2016, 10(4): 24-31.

[18] WANG J W, ZHENG Z Z, WU F, et al. Blockchain-based data market [J]. Big Data, 2020, 6(3): 21-35.

[19] NAKAMOTO S. Bitcoin: A peer-to-peer electronic cash system [J]. Decentralized Business Review, 2008: 21260.

[20] BUTERIN V. A next-generation smart contract and decentralized application platform[J]. White Paper, 2014, 3(37).

[21] ALTAWY R, ELSHEIKH M, YOUSSEF A M, et al. Lelantos: A blockchain-based anonymous physical delivery system[C]// 2017 15th Annual Conference on Privacy, Security and Trust (PST). IEEE, 2017.

[22] CHIESA A, GREEN M, LIU J, et al. Decentralized anonymous micropayments[C]// Annual International Conference on the Theory and Applications of Cryptographic Techniques. Cham: Springer, 2017: 609-642.

[23] KOPP H, MÖDINGER D, HAUCK F, et al. Design of a privacy-preserving decentralized file storage with financial incentives [C]// 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). IEEE, 2017: 14-22.

[24] LEIBA O, YITZCHAK Y, BITTON R, et al. Incentivized delivery network of IoT software updates based on trustless proof-of-distribution[C]// 2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). IEEE, 2018: 29-39.

[25] KIYOMOTO S, RAHMAN M S, BASU A. On blockchain-based anonymized dataset distribution platform[C]// 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA). IEEE, 2017: 85-92.

[26] ZHOU J, TANG F, ZHU H, et al. Distributed data vending on blockchain[C]// 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2018: 1100-1107.

[27] YANG L, JIN R. Distance metric learning: A comprehensive survey[J]. Michigan State Universiy, 2006, 2(2): 4.

[28] JUNG T, LI X Y, HUANG W, et al. Accounttrade: Accountable protocols for big data trading against dishonest consumers[C]// IEEE INFOCOM 2017-IEEE Conference on Computer Communications. IEEE, 2017: 1-9.

[29] DAI W, DAI C, CHOO K K R, et al. SDTE: A secure blockchain-based data trading ecosystem[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 725-737.

[30] MCKEEN F, ALEXANDROVICH I, BERENZON A, et al. Innovative instructions and software model for isolated execution [J]. Hasp@ isca, 2013, 10(1).

[31] XIONG W, XIONG L. Smart contract based data trading mode using blockchain and machine learning[J]. IEEE Access, 2019, 7: 102331-102344.

[32] PARAMESWARAN S, WEINBERGER K Q. Large margin multi-task metric learning[J]. Advances in Neural Information Processing Systems, 2010, 23: 1867-1875.

[33] ZHAO Y, YU Y, LI Y, et al. Machine learning based privacy-preserving fair data trading in big data market[J]. Information Sciences, 2019, 478: 449-460.

[34] POETTERING B, STEBILA D. Double-authentication-preventing signatures[J]. International Journal of Information Security, 2017, 16(1): 1-22.

[35] FERNÁNDEZ-CARAMÉS T M, FRAGA-LAMAS P. A Review on the Use of Blockchain for the Internet of Things[J]. IEEE Access, 2018, 6: 32979-33001.

[36] PATEL V. A framework for secure and decentralized sharing of medical imaging data via blockchain consensus[J]. Health Informatics Journal, 2019, 25(4): 1398-1411.

[37] YU H, ZHANG Z Y, LIU J W. Research on the Expansion Technology of Bitcoin Blockchain [J]. Chinese Journal of Computer Research and Development, 2017, 54(10): 2390.

[38] ZHANG P, HELVIK B E. Towards green P2P: Analysis of energy consumption in P2P and approaches to control[C]// 2012 International Conference on High Performance Computing & Simulation (HPCS). IEEE, 2012: 336-342.

[39] BRUCE J D. The mini-blockchain scheme [EB/OL]. http://cryptonite.info/f i les/mbc-scheme-rev3. pdf.

[40] MEIKLEJOHN S, POMAROLE M, JORDAN G, et al. A fistful of bitcoins: characterizing payments among men with no names [C]// Proceedings of the 2013 Conference on Internet Measurement Conference. 2013: 127-140.

[41] HALPIN H, PIEKARSKA M. Introduction to Security and Privacy on the Blockchain[C]// 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). IEEE, 2017: 1-3.

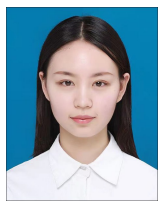
[42] Van Saberhagen N. CryptoNote v 2. 0[EB/OL]. https://www.mendeley.com/catalogue/7e4cdb00-7955-30e1-9185-32a1801bd94b/.

[43] MOORE C, O'NEILL M, O'SULLIVAN E, et al. Practical homomorphic encryption: A survey[C]// 2014 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2014: 2792-2795.

[44] SASSON E B, CHIESA A, GARMAN C, et al. Zerocash: Decentralized anonymous payments from bitcoin [C]//2014 IEEE Symposium on Security and Privacy. IEEE, 2014: 459-474.

[45] LI R, SONG T, MEI B, et al. Blockchain for large-scale internet

of things data storage and protection[J]. IEEE Transactions on Services Computing, 2018, 12(5): 762-771.
[46] AL-RIYAMI S S, PATERSON K G. Certificateless public key cryptography[C]//International Conference on the Theory and Application of Cryptology and Information Security. Berlin: Springer, 2003: 452-473.
[47] OUADDAH A, ABOU ELKALAM A, OUAHMAN A A. Towards a novel privacy-preserving access control model based on blockchain technology in IoT[M]// Europe and MENA Cooperation Advances in Information and Communication Technologies. Cham: Springer, 2017: 523-533.
[48] LIANG X, ZHAO J, SHETTY S, et al. Integrating blockchain for data sharing and collaboration in mobile healthcare applications[C]// 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). IEEE, 2017: 1-5.
[49] MERKLE R. Protocols for Public Key Cryptosystems[C]// Proceedings of the 1980 IEEE Symposium on Security and Privacy (Oakland, CA, USA). 1980.
[50] KAASHOEK M F, KARGER D R. Koorde: A simple degree-optimal distributed hash table[C]// International Workshop on Peer-to-Peer Systems. Berlin: Springer, 2003: 98-107.
[51] YU S, WANG C, REN K, et al. Achieving secure, scalable, and fine-grained data access control in cloud computing[C]// 2010 Proceedings IEEE INFOCOM. IEEE, 2010: 1-9.
[52] BENET J. Ipfs-content addressed, versioned, p2p file system [J]. arXiv: 1407.3561, 2014.



CAO Meng, born in 1997, postgraduate. Her main research interests include blockchain and big data processing and analysis.

SHI Hong-zhou, born in 1971, a senior engineer. His main research interests include blockchain, big data processing and analysis, Internet of things security and other technologies.

(上接第 175 页)

[11] LIU P, QIU X, HUANG X. Recurrent neural network for text classification with multi-task learning[J]. arXiv: 1605.05101, 2016.
[12] KIM Y. Convolutional Neural Networks for Sentence Classification[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Qatar, 2014: 1746-1751.
[13] ZHOU P, QI Z, ZHENG S, et al. Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling[J]. arXiv: 1611.06639, 2016.
[14] XING X, SUN G Z. Dual-channel word vectors based acrnn for text classification. [J/OL]. (2020-12-14)[2021-01-21]. https://doi.org/10.19734/j.issn.1001-3695.
[15] DU L, CAO D, LIN S Y, et al. Extraction and Automatic Classification of TCM Medical Records Based on Attention Mechanism of BERT and Bi-LSTM[J]. Computer Science, 2020, 47(S2): 416-420.
[16] BAI F B, CHANG L, WANG S F, et al. An Improved method study on the extracting keywords in chinese Judgment documents[J]. Computer Engineering and Applications, 2020, 56(23): 153-160.
[17] HOCHSREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
[18] DONG Y R, LIU P Y, LIU W F, et al. A text classification model based on BiLSTM and label embedding[J]. Journal of Shandong University(Natural Science), 2020, 55(11): 78-86.
[19] SUN H, CHEN Y Q. Chinese text classification based on BERT and attention. [J/OL]. (2021-01-06) [2021-01-21]. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=XXWX2021010500E&DbName=CAPJ2021.
[20] WANG H T, SONG W, WANG H. Text classification method based on hybrid model of LSTM and CNN[J]. Journal of Chinese Computer Systems, 2020, 41(6): 1163-1168.
[21] WANG G S, HUANG X J. convolution neural network text classification model based on Word2vec and improved TF-IDF [J]. Journal of Chinese Computer Systems, 2019, 40(5): 1120-1126.
[22] LI Y H, LIANG S C, REN J, et al. Text classification method based on recurrent neural network variants and convolutional neural network[J]. Journal of Northwest University(Natural Science Edition), 2019, 49(4): 573-579.

JING Li, born in 1971, Ph.D, professor, is a member of China Computer Federation. Her main research interests include artificial intelligence and information security.

HE Ting-ting, born in 1996, postgraduate. Her main research interests include natural language processing and data mining.