计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210600015

基于区块链的云存储安全研究进展

徐 堃 付印金 陈卫卫 张亚男
陆军工程大学指挥控制工程学院 南京 210007
(1255284410@qq.com)

摘 要 云存储使得用户能够随时随地通过网络连接按需获取廉价的在线存储服务,但因云服务提供商、第三方机构和用户的不可信以及不可避免的恶意攻击,存在诸多云存储安全漏洞。区块链拥有去中心化、持久性、匿名性和可审计性的特点,具有建立可信平台的潜力。因此,基于区块链技术的云存储安全机制研究已成为一种研究趋势。据此,首先概述云存储系统安全架构与区块链技术的安全性,然后从访问控制、完整性验证、重复数据删除和数据溯源4个方面进行文献综述与对比分析,最后对基于区块链的云存储安全进行技术挑战分析,并总结全文与展望未来。

关键词:云存储安全;区块链;访问控制;完整性验证;重复数据删除;数据溯源
中图法分类号 TP311

Research Progress on Blockchain-based Cloud Storage Security Mechanism

XU Kun,FU Yin-jin,CHEN Wei-wei and ZHANG Ya-nan
College of Command and Control Engineering,Army Engineering University of PLA,Nanjing 210007,China

Abstract Cloud storage enables users to obtain cheap online storage services on demand through network connection anytime and anywhere. However,due to the untrustability of cloud service providers,third-party institutions and users as well as the inevitable malicious attacks,there are many security vulnerabilities of cloud storage. Blockchain has the potential to build a trusted platform with its characteristics of decentralization,persistence,anonymity and auditability. Therefore,the research on cloud storage security mechanism based on blockchain technology has become a research trend. Based on this,the security architecture of cloud storage system and the security of blockchain technology are first outlined,then the literature review and comparative analysis are conducted from four aspects of access control,integrity verification,data deduplication and data provenance. Finally,the technical challenges of blockchain-based cloud storage security mechanism are analyzed,summarized and prospected.

Keywords Cloud storage security,Blockchain,Access control,Integrity verification,Data deduplication,Data provenance

1 引言

近年来,云存储在数据存储量、安全性、可靠性、成本等方面的优势使其逐步取代了传统的存储方式。用户将数据外包给云服务提供商(Cloud Service Provider,CSP),并通过互联网远程访问数据。CSP为用户提供了一种更为高效、灵活的数据管理方式,使得用户无需关注基础设施的维护,就能用相对低廉的成本换得由服务等级协议(Service-Level Agreement,SLA)保证的较高的服务水平。

尽管云存储系统有一套保证安全性的机制,如访问控制、完整性验证等,但由于CSP是"半诚实的"[1],许多服务都是由第三方机构参与的,因此用户、CSP、第三方机构之间无法完全建立信任,且恶意攻击不可避免,云存储仍面临多方面的安全挑战。CIA是信息系统安全的3个要素,分别为机密性(Confidentiality)、完整性(Integrity)和可用性(Availability)[2]。在机密性方面,恶意用户可能通过勾结、攻击等方式干涉云访问控制流程以非法访问数据,用户的隐私也可能在数据挖掘的过程中被泄露;在完整性方面,云中的数据不受数据所有者的掌控,无论是原始数据,还是软件在执行过程中产生的中间数据,其完整性都面临被窃取、篡改、删除以及服务中断的威胁;在可用性方面,数据可能因为软硬件损坏、带宽不足或外部力量而不可用,但CSP可能无法找到数据不可用的原因而无法及时恢复数据,云存储系统也可能因为冗余数据过多而使得可用性降低。

区块链技术是一种新兴技术,它不仅颠覆了以银行为中心的传统互联网交易方式,在其他应用场景中也具有广泛的应用前景,如跨境电商、产权保护、征信平台、互联网医疗等。区块链具有去中心化、持久性、匿名性和可审计性的特点,能

到稿日期:2021-06-01 返修日期:2021-07-12
基金项目:国家自然科学基金(61402518);江苏省自然科学基金(BK20191327)
This work was supported by the National Natural Science Foundation of China(61402518) and Natural Science Foundation of Jiangsu Province (BK20191327).
通信作者:陈卫卫(njcww@qq.com)

够建立可信平台，且智能合约能实现流程的自动控制，这表明区块链具有在安全云存储系统中增强数据保护机制的潜力。例如，将访问控制数据存储在区块链上，将访问控制流程以智能合约的方式执行可以避免他人对访问控制过程的干涉，可使其透明、正确地运行；通过在区块链上保存元数据，云中的原始数据与软件中间数据都可借助溯源链来验证其完整性；将数据溯源信息保存在区块链中可以准确定位数据异常；可以借助分布式的区块链来管理多云，完成多云重复数据删除以减少冗余等。

将区块链技术与云存储安全结合已成为一种研究趋势，然而区块链技术在云存储安全中的研究还处于初期阶段，集中化管理的传统云存储模式与分散化的区块链技术相结合仍缺乏技术经验。为了促进基于区块链的云存储安全研究，本文从不同方面对研究工作进行了整理，首先介绍了云存储安全与区块链技术的安全性，其次从访问控制、完整性验证、重复数据删除、数据溯源 4 个方面，重点阐述了区块链在云存储安全中的研究进展，然后分析了基于区块链的云存储安全的技术挑战，最后总结全文并展望未来。

2　相关工作

2.1　云存储安全概述

云存储是在云计算发展历程中孕育出的一种新的概念，它通过集群应用、网格技术、分布式文件系统等技术，将网络中的大量存储设备通过软件方式协同起来，为用户提供数据存储和业务访问接口[3]。与传统的存储方式相比，云存储具有存储空间海量、安全策略统一部署、可靠性强、成本低廉、无需关注基础设施建设维护的特点。安全的云存储系统具有数据加密、数据分片、数据容灾备份、访问控制、完整性验证、重复数据删除、数据溯源、数据确定性删除、密文搜索等机制。云存储系统的安全架构如图 1 所示，自上而下包含访问层、接口层、基础管理层和存储层[4]。

在访问层，用户通过接口层与 CSP 进行交互。数据所有者将数据加密后，通过接口层将数据通过网络传输至基础管理层，数据所有者通过访问控制接口部署访问控制策略，用户依据访问控制策略获权访问数据。此外，访问层还调用接口层的数据完整性验证接口。接口层是访问层与基础管理层的交互工具。基础管理层是系统最核心的部分，通过集群、分布式文件系统、网格计算等技术实现存储资源的虚拟化，为用户提供存储服务。它负责收集日志，管理元数据，建立数据索引，进行数据溯源，对上层提供数据完整性验证、访问控制、密文搜索等服务，提升用户体验；它还指导下层的重复数据删除、数据容灾备份、数据分块存储，以提高系统自身的稳定性。存储层是存储数据的物理层，数据密文以及附加信息（例如元数据）均存储在该层，存储层设有存储设备管理系统，能够实现物理链路管理及硬件设施的监控和维护。

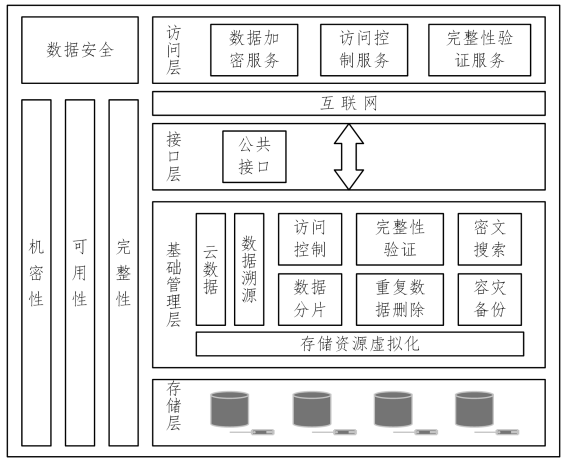

图 1　云存储系统的安全架构

Fig. 1　Security architecture of cloud storage system

传统云存储安全机制存在诸多痛点，本文针对访问控制、完整性验证、重复数据删除及数据溯源 4 个方面的痛点进行分析，并对基于区块链的解决方案进行整理，具体分析见第 3 节。

2.2　区块链安全性概述

自 2008 年比特币（bitcoin）[5]诞生后，其背后的区块链技术作为一项在安全领域的创新技术，受到了全世界开发者的瞩目。区块链可以被看作一个开放的分布式账本，不断有新的交易被打包到新的区块中，随着区块的不断更新，区块链会按照时间有序增长。如图 2 所示，区块中包含了存储交易的 Merkle 哈希树（MHT）、块哈希、父哈希、版本号、随机数、时间戳等信息，其中块哈希是根据区块中其他信息生成的，包括父哈希。因此，如果要更改区块链中的一个区块，则该区块之后的所有区块都必须更改才能不违背区块链的构成规则。

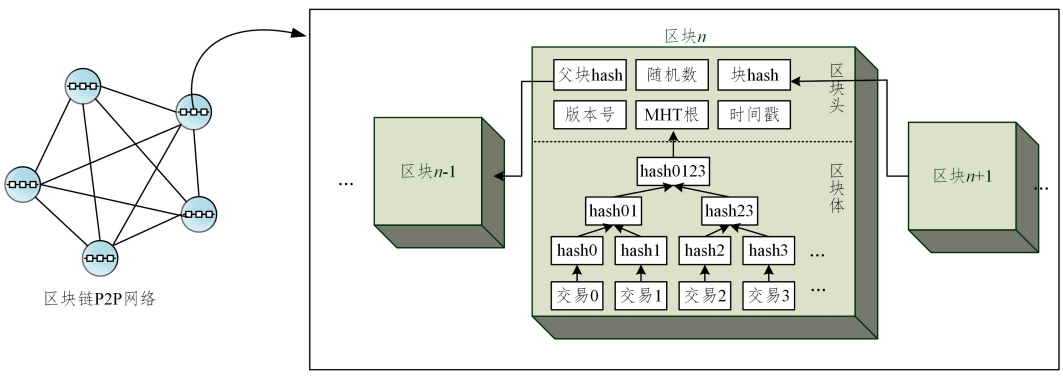

图 2　区块链的结构

Fig. 2　Structure of blockchain

区块链技术是多种计算机技术的新型应用模式，它通过集成一些核心技术，如加密哈希、数字签名、共识机制、时间戳技术等，能够在节点不互相信任的情况下，实现分布式系统中的去中心化点对点信息传输。近年来，随着区块链技术的热度飙升，各个领域都开展了对区块链的研究，如物联网领域、云计算领域、金融领域、医疗保健、电子政务等。总的来说，区块链具有以下特性[6]。

(1)去中心化。在传统中心化的交易系统中，任何一笔交易都必须经过中心节点的验证，增加了中心服务器的工作负载，且具有中心服务器瘫痪导致服务不可用的风险。区块链网络中，多个参与方通过 P2P 传输和共识机制共同维护账本。

(2)持久性。一个交易一旦被添加到一个有效的区块链中，就几乎没有被删除或者修改的可能。

(3)匿名性。在区块链网络中，用户使用系统生成的地址(由用户公钥的 hash 生成)进行交易，用户的真实信息可以被隐藏。

(4)可审计性。正如比特币通过 Unspent Transaction Output(UTXO)模型来存储用户余额，每次的输入都是由上次的输出分割而来[7]。区块链中的每笔交易都会引用以前的交易状态，一旦交易添加到区块链上，交易状态就会更新。因此，任何交易记录都可以被轻松地追溯和审查。

区块链按照其发展历程可分为 1.0，2.0，3.0 这 3 个发展阶段。1.0 阶段的产物是以 Bitcoin 为代表的数字货币。2.0 阶段是在数字货币的基础上，发展成为以智能合约的应用为代表的可编程区块链。智能合约可以自动化执行合约流程，无须第三方中心机构的介入。将智能合约以数字化的形式写入区块链，可以保证其存储、读取、执行的过程可跟踪和不可篡改，代表平台为 Ethereum。3.0 阶段为金融领域之外的各行各业提供基于区块链的去中心化解决方案，代表产物为 Hyperledger Fabric，EOS。

将区块链技术与云存储相融合，能够使其在性能增强上发挥出巨大潜力。目前在基于区块链的云存储项目中，具有代表性的有 Filecoin，Storj，Sia，Lambda 等。Filecoin 是一个在区块链上运行的分布式平台[8]，以 IPFS[9] 为基础设施层，通过经济激励促进 IPFS 的发展，该平台中矿工提供磁盘空间和带宽，通过向客户提供存储空间来获得加密货币 FIL 奖励。Storj[10] 是一个基于区块链的端对端分布式云存储平台，该平台通过加密使用户通过私钥以安全去中心化的方式管理数据。Sia[11] 通过纠删码、加密技术和区块链技术，在传统云存储的基础上改善了安全隐私方面的问题，通过为提供存储空间的用户发放 Siacoin(云储币)来激励更多拥有闲散空间的用户成为存储供应商。Lambda[12] 是一个以 Ethereum 为基础链建立的去中心化存储网络，在 Lambda Chain 共识网络上实现数据存储、完整性和安全性验证以及市场运行维护，具有安全可靠、存储空间可无限扩展的特点，旨在为下一代互联网提供基础设施，推动互联网的去中心化发展。

3 基于区块链的云存储安全

3.1 基于区块链的云存储访问控制

访问控制指保护数据免受用户对其进行未经授权的访问和操作。传统云访问控制模型包括访问控制列表(Access Control List，ACL)、基于任务的访问控制(Task-Based Access Control，TBAC)、基于属性的访问控制(Attribute-Based Access Control，ABAC)、基于使用控制的访问控制(UCON-Based Access Control)、基于 Bell-LaPadula 的访问控制(BLP-Based Access Control)等。ACL 即为每个资源维护一个访问控制列表。在 TBAC 中，数据用户对数据对象的访问权限随着任务执行过程中的上下文环境的变化而变化，它对不同工作流或者同一工作流的不同任务实例有着不同的访问控制策略，且具有时效性。ABAC 则根据主体(请求者)、对象、环境的属性是否满足策略来决定请求者是否具有访问权限。对于基于 UCON 的访问控制，文献[13]提出了一个 UCON 模型来解决云环境中的主体属性可变性和义务处理。对于基于 BLP 的访问控制，文献[14]提出了一种基于 BLP 模型的虚拟机系统，实现了虚拟机隔离和共享。

为了保护数据隐私，防止云服务提供商窃取数据，大部分数据在被加密后才能被上传到云存储服务器中，基于属性的加密(Attribute-Based Encryption，ABE)技术可解决加密数据共享问题。目前，基于密文策略的属性加密(Ciphertext-Policy ABE，CP-ABE)[15] 是云存储中提供安全访问控制机制的最合适的技术之一。CP-ABE 将策略嵌入密文中，数据所有者通过设计策略来决定拥有哪些属性的人能够访问这份密文，做到云存储细粒度访问控制，将属性嵌入到密钥中，当且仅当数据用户的属性满足数据所有者设定的策略时，数据用户才可以解密。

由于访问控制策略执行所需的存储和计算量庞大，这些云访问控制方案都通过一个集中的服务器来验证访问权限，传统的云访问控制方案需要一个中心化服务器存储访问权限，基于加密的访问控制需要一个或多个中心化服务器分发密钥。

如图 3 所示，云访问控制模型可以抽象为数据所有者、数据用户、云服务提供商、访问控制服务器、系统管理员 5 个实体。

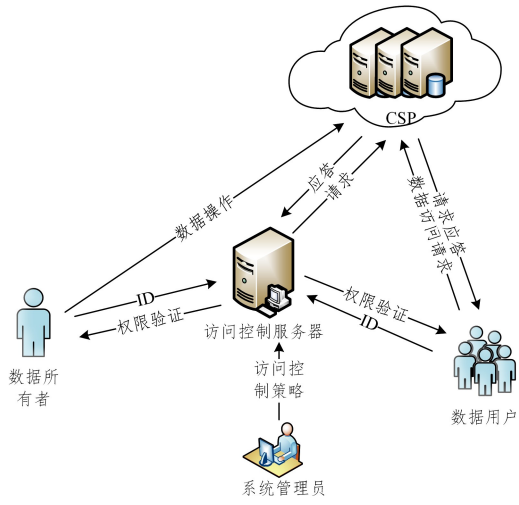

图 3 传统云访问控制模型

Fig. 3 Traditional cloud access control framework

中心化的服务器可能会恶意地泄露用户数据隐私、私自篡改数据的访问权限列表，或者给没有权限的用户“开后门”。此外，中心化的服务器也可能因为故障或黑客的恶意攻击而损害访问控制信息的可用性和完整性。

借助于区块链的只增不减、不可篡改的持久性特点，将访问控制策略直接存储在区块链中可以防止策略被修改或删除。利用智能合约进行权限验证可以防止用户越权行为。此外，还可将访问控制策略以程序化的形式写入智能合约，由智能合约根据条件自动生成访问控制策略，避免了恶意的人为干预。基于区块链的云访问控制模型大致如图 4 所示，访问控制策略既可以由数据所有者部署至区块链，也可以由访问控制系统管理员根据数据所有者制定的访问控制规则生成并部署至区块链，亦或由智能合约自动生成。通过调用区块链接口可获取数据用户的操作权限，用户是否具有权限进行相应操作可由智能合约验证，也可由 CSP 验证，CSP 据此响应用户的操作请求。区块链也能以日志的形式记录数据的访问历史，以便追溯异常行为。

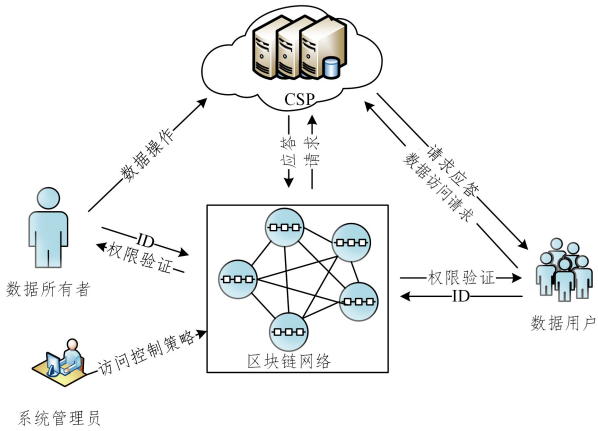

图 4　基于区块链的云访问控制模型

Fig. 4　Blockchain-based cloud access control framework

在 ACL 模型中，文献[16]提出了基于区块链的云访问控制机制 BACC，其中 ACL 由数据所有者更新，当云服务器要授权用户并验证用户权限时，向区块链网络发送事务请求，由智能合约进行验证并响应，因此云服务器无法控制访问权限列表和授权过程。在此工作的基础上，其还提出使用 Shamir 秘密共享方案，将密钥分成多块存储在区块链网络的矿工节点中，但区块链的节点不是可信节点，仍无法保证密钥的安全。此外，该方案中的数据所有者、数据用户、云存储服务器需要额外安装 Ethereum 客户端，才能与区块链网络进行通信，以调用智能合约实现访问控制。该模型包含 4 个智能合约，其中，ACP 合约维护每个数据对象的用户列表和授予他们权限并提供更新、验证权限的接口；AUD 合约维护审计信息列表；LKP 合约管理主节点（即存储密钥的矿工节点）；CLP 合约维护智能合同及其地址的列表。文献[17]也做了类似的工作，与 BACC[16] 不同的是，文献[17]将密钥存储在基于安全假设的安全服务器网络中，它的体系结构更复杂，授予权限需要更长的时间。文献[18]在 EOS 中实现了 Auth-PrivacyChain，设计了访问控制、授权和授权撤销的流程，并在区块链中加密访问控制信息。

文献[19]针对不可信云环境中的访问控制，在无需动态改变密钥的情况下提出了一个基于区块链的动态 CP-ABE 模型，将密钥生成、访问控制策略、权限更改或撤销、数据访问请求等过程以日志形式存储在区块链中，在 Ethereum 中进行测试。该方案中的大多数功能在客户端实现，在 EVM 中执行两种智能合约，其中 CA 合约初始化系统，AA 合约管理属性。

文献[20]针对密钥生成器（PKG）可能带来密钥滥用和泄露的问题，提出了一种基于 Ethereum 区块链的方案，使密钥分发不依赖中心化服务器来实现。此外，数据所有者可以为数据用户附加一个有效的访问周期，只有在有效访问期内，并且数据用户的属性集满足访问控制策略时，数据用户才能正确执行数据解密算法。但该方案中数据用户每次申请访问时，需要与数据所有者进行通信，这增加了授予权限所需的时间。

文献[21]针对基于多授权中心的 CP-ABE 访问控制方案的单点故障和通信开销大的问题，提出了基于区块链的多授权中心访问控制方案 BMAC，引入 Shamir 秘密共享方案和 Hyperledger Fabric 来实现每个属性由多个机构联合管理，避免单点故障，并利用区块链技术在多个机构之间建立信任，利用智能合约为跨多个管理域管理的属性计算令牌，从而减少数据用户端的通信和计算开销。此外，区块链以安全和可审计的方式记录了访问控制过程。由于区块链交易具有透明性，该文献还考虑了融合链上链下存储来保证数据的安全性。

文献[22]针对 CP-ABE 中间实体造成的信任成本高、单点故障等问题，提出了基于区块链和 CP-ABE 的访问控制方案 TrustAccess，并在 java 环境中进行了实验评估。一方面，其提出 HP-CP-ABE 方案，在满足大规模访问需求的同时保证策略隐私；另一方面，其使用 ElGamal 同态加密来确保授权验证期间的属性隐私。在 TrustAccess 中，只有数据用户、数据所有者和区块链参与访问控制，没有其他中间实体的参与。但 TrustAccess 基于合数阶群实现，效率低下。

文献[23]针对现有采用 ABE 的访问控制方案效率和灵活性低下的问题，提出了一种快速可追踪的基于属性的动态访问控制加密方案 TABE-DAC，在区块链上进行隐私保护和细粒度数据共享，大大提高了 ABE 的计算效率和存储成本。TABE-DAC 不仅支持追踪泄露私钥的恶意用户，防止密钥的非法共享和滥用，还支持访问控制策略的灵活更新。TABE-DAC 与 TrustAccess 的实验环境相同，基于素数阶群实现，效率显著优于 TrustAccess。

然而，ABE 基于双线性对实现，双线性对操作成本昂贵，不适用于物联网环境中大量计算资源有限的设备，因此现有方案将该操作外包给云，但云是不可信的，用户仍需消耗计算资源以验证结果的正确性。文献[24]提出了一种基于 ABE

和区块链技术的轻量级解密访问控制方案 LBAC，借助于 Hyperledger Fabric 联盟链以保证外包计算的正确性。用户的访问行为也被记录在区块链上，以增加数据访问的透明度。此外，为了激励用户参与区块链系统，LBAC 设计了用户可信度激励机制，用户访问失败的次数越多，可信度越低，其背书节点就越少。

文献[25]还提出了一种新型服务模式 BEAAS（Blockchain Enabled Attribute-Based Access Control as a Service），该模式将基于区块链的 ABAC 作为一种服务。对各种 ABAC 组件和访问历史的所有更改都会通过智能合约添加到 Ethereum 中。通过 BEAAS，用户能够验证访问控制决策是否正确执行。在未来，BEAAS 计划实现支持动态访问控制和从区块链信息中重建 ABAC 模式的弹性功能。

文献[26]针对传统 ABE 的 PKG 不可信及集中式云存储架构的单点故障和不诚实问题，提出了一种将分布式存储系统 IPFS、Ethereum 区块链和 ABE 相结合的数据存储和共享方案。该方案中，数据所有者依据区块链数据为数据用户分配密钥，同时通过智能合约实现了对分散存储系统密文的关键字搜索功能。然而，该方案没有实现用户属性撤销和访问策略更新的功能。

此外，FaaS（Federation-as-a-Service）[27]是一种新的云协作模式，在该模式下，可以跨越私有云基础架构共享数据。文献[28]针对联合云环境中的访问控制系统不能保证在处理请求时不被绕过的问题，提出了一个基于 Ethereum 的去中心化实时监控架构 DRAMS，该架构可以在特定威胁模型的假设下检测到违规操作记录。DRAMS 使用 ABAC 模型，代理收集日志并将其上传到区块链网络，分析员根据日志分析访问控制决策的正确性。文献[29]提出了一种新的基于 Ethereum 的联合云身份和访问管理系统。该系统对数据实施 ABE 策略，使用区块链技术和英特尔 SGX 可信硬件来保证策略评估过程的完整性，使用区块链来确保用户身份属性和访问控制策略不会被恶意用户修改，以及英特尔 SGX 保护策略实施过程的完整性和机密性。

表 1 列出了从访问控制模型、是否为动态访问控制、系统组成部分、云服务架构、区块链平台等方面对上述基于区块链的云存储访问控制方案进行对比分析的结果。动态 CP-ABE 模型相比于静态的 ACL 访问控制模型更能满足云存储中的细粒度访问控制需求，系统中除了 CSP、数据用户和区块链网络之外的其他网络或者其他第三方实体会增加系统的不安全因素，与区块链 2.0 阶段的以太坊平台相比，区块链 3.0 阶段的平台吞吐量与云存储服务中庞大的数据量更加匹配。因此，未来的基于区块链的云存储访问控制应侧重于构建动态的基于 CP-ABE 的无其他可信中心的系统。此外，表 1 还对比了云服务架构，基于区块链的云存储访问控制在主流的基于 C/S 架构的云存储、分布式云存储以及 FaaS 多云存储中都能发挥作用。

表 1 基于区块链的云存储访问控制方案对比分析

Table 1 Comparisons of blockchain-based cloud access control framework

	访问控制模型	动态/静态	系统组成部分	云服务架构	区块链平台
BACC[16]	ACL	静态	CSP,DO,DU,BCN	C/S	Ethereum
文献[17]	ACL	静态	CSP,DO,DU,BCN,安全服务器网络	C/S	N/A
AuthPrivacyChain[18]	ACL	动态	CSP,DO,DU,BCN	C/S	EOS
文献[19]	CP-ABE	动态	CSP,客户端,BCN(EVM),CA,AA	C/S	Ethereum
文献[20]	CP-ABE	静态	CSP,DO,DU,BCN	C/S	Ethereum
BMAC[21]	多授权中心 CP-ABE	静态	CSP,DO,DU,BCN,CA,AAs	C/S	Hyperledger Fabric
TrustAccess[22]	CP-ABE	静态	CSP,DO,DU,BCN	C/S	Java 创建
TABE-DAC[23]	ABE	动态	CSP,DO,DU,BCN,CA	C/S	Java 创建
LBAC[24]	ABE	静态	CSP,DO,DU,BCN,CA,AAs	C/S	Hyperledger Fabric
BEAAS[25]	ABE	静态	CSP,DO,DU,BCN,APIs	C/S	Ethereum
文献[26]	ABE	静态	IPFS,DO,DU,BCN	P2P	Ethereum
DRAMS[28]	ABAC	静态	CSPs,BCN,探测代理,分析员	FaaS C/S	Ethereum
文献[29]	ABE	静态	CSPs,BCN,DU,链外存储,身份管理器,访问控制管理器	FaaS C/S	Ethereum

注：DO 为数据所有者；DU 为数据用户；BCN 为区块链网络；CA 为证书颁发机构；AA 为属性权威机构

3.2 基于区块链的云数据安全完整性验证

完整性验证指用算法验证数据对象是否被修改。根据完整性验证方法的不同，Priyadharshini 在文献[30]中对传统云端数据完整性验证方法进行了分类总结，其可以大致分为以下几类：基于整体数据依赖标签、基于数据块依赖标签、基于数据独立标签和基于数据复制的验证协议。然而，传统云数据完整性控制方案的基本思路是在云端与用户或者云端、用户与第三方机构（公共审计）之间构造协议，通过“挑战-应答”的方式进行。由于完整性验证的计算量较大，用户在本地验证的成本过高、资源不足，且 CSP 和用户之间互不信任，用户可能会伪造数据完整性受损的假信息骗取 CSP 的赔偿费用等问题，因此用户本地验证逐渐被公共审计取代。如图 5 所示，公共审计中，用户委托一个第三方审计机构（TPA）完成完整性验证工作，然而 TPA 也可能不诚实[31]。例如，公共审计过程是周期性进行的，TPA 为了避免成本，可能会在一次没有问题的验证之后，不执行接下来若干次验证过程，但是会生成一份好的报告；TPA 可能与 CSP 进行合谋，如共同对用户隐瞒数据损坏行为，或者只验证完整性好的数据块；TPA 并不由用户控制，可能会泄露数据等。区块链技术的去中心化架构可以使数据的审计摆脱对可信第三方的依赖，将区块链引入云间数据迁移能够有效验证数据的完整性。

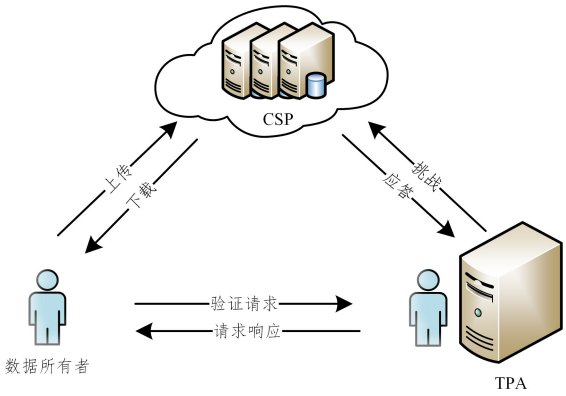

图 5　公共审计模型

Fig. 5　Public audit framework

区块链可用于加强云数据的完整性。文献[32]提出了一个将区块链技术用于加强 FaaS[27] 中云数据完整性的初步构想，由于区块链的低吞吐量、高延迟和弱稳定性，设计方案中有两层区块链，第一层用于确保性能，实现低延迟和高吞吐量，第二层用于确保完整性，但性能较差。两个层之间进行交互以获得整体性能的提高和对数据完整性的有效保证。

部分研究直接通过对比原始数据 hash 来验证完整性。文献[33]提出了一种名为 Zeppar 的区块链私链模型，将文件 hash 值上传至区块链中，通过对比 hash 值来判断数据是否完整。文献[34]利用移动代理技术，在云中部署分布式虚拟机代理模型，并在其之上建立区块链网络，智能合约根据 MHT 生成的文件所对应的唯一哈希值来监控数据变化。

另一部分研究通过生成挑战信息来验证完整性。文献[35]提出了在 Bitcoin 公链上实现的基于身份的云数据公共审计方案 IBPA。它包含 PKG、TPA、CSP 和客户端，用区块链的随机数构造不可预测且易于验证的挑战消息，从而防止恶意的 TPA 伪造审计结果来欺骗用户。这种方法的局限性在于，TPA 仍具有安全问题以及更高的计算开销和通信开销。

文献[36]用区块链的最新区块 hash 构造挑战信息以确保随机性，TPA 将审计记录发送至区块链上。针对每个数据，将所有审计信息串联起来构成每个数据的审计链，用户只需验证审计链的最新一个审计记录的有效性，而无需验证 TPA 的所有审计行为，就能确保审计的有效性。

与上述有 TPA 的方案不同，文献[37]提出了一个基于区块链的无可信第三方的公共审计方案框架，用户首先用 KGC 颁发的私钥计算数据的标签，再将原数据与相应的标签上传至 CSP。当用户有完整性验证需求时，同时向 CSP 与区块链节点发送请求和挑战信息，CSP 将应答信息返回给区块链节点，由智能合约完成验证过程，并将结果保存至区块链中。此外，作者分别针对安全和高效两种需求设计了两种共识机制。实验证明，该机制的性能优于 IBPA。

文献[38]提出的架构也完全省去了第三方审计机构，只有互不信任的数据所有者的 CSP。数据所有者将轻量级验证标签存储在区块链上，并通过构建 MHT 来证明云数据的完整性。具体来说，数据所有者首先将文件分块加密，并通过哈希运算生成加密块的标签并将其存于区块链中，再将加密块上传到 CSP。在审计阶段，区块链网络收到 DO 的请求后，计算指定数据的所有块标签的 MHT 根，CSP 收到 DO 挑战后也计算出所有块标签的 MHT 根，DO 通过对比两者来验证数据的完整性。此外，该方案不需要密钥生成中心，密钥对由数据所有者自己生成。但这种方案存在漏洞，即 CSP 提供的 MHT 根不一定是实时的，有可能是 CSP 在数据刚上传时就生成并保存的，而后可能已经将数据篡改，但是返回最初生成的 MHT 根。

为了避免“把鸡蛋放在同一个篮子里”，许多用户选择将数据放在多个网络存储服务中协同存储。尽管这种存储策略在一定程度上降低了数据完整性的安全风险，提高了效率，但仍缺乏有效的措施确保多云系统下的数据完整性。

文献[39]在多云存储服务环境中，使用用户提供的信息生成随机挑战，使用智能合约对 CSP 发出的挑战应答信息进行审计。除了用于审计的存储监管合约外，该方案还部署了信任管理合约，用于评估每个云存储服务的信任级别。

文献[40]提出了一个基于区块链的多云存储数据审计方案，以保护数据完整性和准确仲裁服务纠纷。用户将其数据外包给多个 CSP，然后和 CSP 共同生成用于数据审计的完整性元数据。在审计阶段，用户随机生成质询随机数，并要求 CSP 根据指定的数据块分别生成完整性证明作为响应。其中，引入区块链来记录数据审计过程中的交互数据，利用同态可验证标签实现了无需 TPA 的低成本批量验证，并利用智能合约来检测不诚实的 CSP。

除了基于客户端-服务器模式(C/S)的传统云存储之外，P2P 云存储因其能够充分利用大量的空闲磁盘空间和提供低廉的存储服务而逐渐兴起。在 P2P 云存储中，每个用户既可以是购买存储服务的客户，也可以是出租存储空间的服务方。

P2P 云存储平台 Sia[11] 用区块链记录相关信息以进行完整性验证，但 Sia 没有提供完整的验证机制，也没有合理选择部分数据块进行验证以提高效率。在资源有限和实时性要求较高的情况下，应制定合理的采样策略完成数据的完整性验证。文献[41]提出了一个基于 P2P 云存储的数据完整性验证框架，并设计了合理的采样策略，使采样验证更加有效。其中，完整性验证依靠 MHT，用户首先用数据块和随机挑战数构造 MHT，将树根 root1 存储至区块链中，再将 MHT 的公共部分和数据块上传至 CSP 中。当需要验证 shard i 的完整性时，用户发送挑战数 r_i 给 CSP，CSP 根据挑战数和原始数据块生成 Digest i，并返回 Digest i 和 MHT 伴随路径给智能合约，智能合约根据这两个参数计算出 root2 并与 root1 进行对比来进行可信的完整性验证。但此方案的验证次数受到挑战数的限制(CSP 在每次验证时可以保存数据块对应的挑战数)。

文献[42]构建了一种基于 Ethereum 的 P2P 云存储方案 DStore，在 DStore 中，任何节点都能自由加入与退出网络，出租本地空余存储资源或者租用他人的存储资源。DStore 运用区块链对 P2P 网络中的节点进行管理，克服了 P2P 网络存

储服务依赖第三方进行公平支付的困难，与文献[41]类似，利用 MHT 通过智能合约而不是第三方机构对数据进行完整性验证，存储日志记录对用户行为进行审计。

云存储中不仅原始数据的完整性至关重要，云日志的完整性也是分析和跟踪云系统中的安全威胁的重要前提。CSP 将一切用户行为记录在日志中，并从中提取用户行为特征和检测非法行为。文献[43]提出了一种基于区块链的日志完整性验证方案，TPA 向 CSP 发送块号和随机数作为挑战信息，CSP 使用固定算法算出应答消息，TPA 将 CSP 应答与从区块链获取的信息进行对比。

拖延审计也是 TPA 不诚实行为的一种。完整性验证的目标之一是尽快发现数据损坏以及时恢复数据，但一些不负责任的公共审计员会因自身网络故障、系统错误或者与不诚实的 CSP 合谋拖延审计并隐瞒拖延行为。文献[44]针对拖延审计提出一种运用区块链技术的无证书公开验证方案 CPVPA，要求审计员将每一次验证结果作为交易打上时间戳存储在区块链中，使用户能够通过检查日志文件的有效性和正确性来核查审计员的行为。此方案包含一个 TPA 和一个 KGC。

此外，研究者也对完整性验证不通过之后的仲裁过程进行了研究。如果 CSP 违反了服务等级协议(Service-Level Agreement，SLA)，则需要根据仲裁协议接受处罚。然而，现有的数据完整性验证方案只关注验证过程如何实现，很少考虑公平仲裁。现有的对于违反 SLA 的处罚都是基于信用的，一旦 CSP 没有按照 SLA 的要求达到相应的服务水平，则需接受处罚，但如果 CSP 拒绝接受处罚，数据所有者则无法获得相应的赔偿，并且维权是一个复杂繁琐的过程，这对数据所有者是不公平的。文献[45]提出了一种基于智能合约的自动公平仲裁方法，将仲裁协议写入区块链智能合约中自动执行，无需任何第三方。该方案中，用户和 CSP 分别接受一定金额的存款作为押金输入，如果 CSP 能通过完整性验证，那么智能合约会将用户的押金作为审计费发送给矿工，并将 CSP 的押金返回给 CSP。如果 CSP 破坏了用户数据的完整性，智能合约将收取 CSP 的押金并补偿用户。

表 2 列出了从完整性验证方法、验证对象、云服务架构、有无 TPA 方面对上述基于区块链的云数据安全完整性验证方案进行对比分析的结果。

表 2 基于区块链的云数据安全完整性验证方案对比分析

Table 2 Comparisons of blockchain-based cloud integrity verification

	验证方法	验证对象	云服务架构	有无 TPA
Zeppar[33]	hash	原始数据	C/S	无
文献[34]	hash	原始数据	C/S	无
IBPA[35]	随机数	原始数据	C/S	有
文献[36]	随机数	原始数据	C/S	有
文献[37]	随机数	原始数据	C/S	无
文献[38]	MHT	原始数据	C/S	无
文献[39]	随机数	原始数据	多云 C/S	无
文献[40]	随机数	原始数据	多云 C/S	无
文献[41]	MHT	原始数据	P2P	无
DStore[42]	MHT	原始数据	P2P	无
文献[43]	随机数	云日志	C/S	有
CPVPA[44]	N/A	审计行为	C/S	有

基于区块链的云数据安全完整性验证能应用在不同的云存储架构中，也能应用于不同的数据验证场景，以提高系统的安全性。相比更加简单的借助 Hash 和随机数的验证方法，基于 MHT 的方法仅用一个 MHT 根就能验证一个数据所有数据块的完整性，减少了所需的区块链交易数量，但基于 MHT 的方法仍存在验证次数受限或者有安全漏洞等问题。无 TPA 的方案能够使系统更加安全，但这可能会增加区块链的计算负担，使得系统与区块链的交互更为频繁。

3.3 基于区块链的云存储重复数据删除

重复数据删除指将重复的数据块用指示符取代从而缩减数据，为存储系统释放更多的存储空间。数据的快速增长给云存储空间和传输带宽造成了很大压力，研究表明，目前 75%的数字数据是冗余的，有效的重复数据删除技术可以节省过半的存储空间。重复数据删除根据执行位置不同分为源端重复数据删除和目的端重复数据删除。源端重复数据删除指数据在到达目标服务器之前进行重复删除处理，然而目的端重复数据删除指数据在到达目标服务器之后再进行重复删除处理。

重复数据删除依然依赖不可靠的实体进行重复数据检测。重复数据删除与传统的加密算法不能一起工作，不同用户用不同私钥加密数据，会使得相同数据生成不同密文，导致重复数据删除不可行。为了解决这个问题，文献[46]提出了收敛加密(Convergent Encryption，CE)，用户通过计算数据的哈希值来生成收敛密钥(CK)以加密数据。因此，不同的用户将为 CE 中相同的数据生成相同的密文，从而使加密数据的重复数据删除变得可行。然而，当数据可预测时，CE 算法可以被暴力破解。文献[47]提出了消息锁加密算法(Message Locked Encryption，MLE)，CE 是该算法的一种，但 MLE 算法仍不能提供语义安全性。此外，以往的工作通常引入一个密钥管理服务器来管理密钥，由于大多数支持密钥管理的重复数据删除方案都过度依赖密钥管理服务器，因此容易受到共谋攻击，导致密钥泄露。因此，抵御共谋攻击，确保密钥的安全是一个巨大的挑战。再者，在重复数据删除的过程中，由于存储服务器中通常只保留一个副本，数据的完整性极易受到恶意攻击者和服务中断的威胁，因此进行数据审计以及在可靠性与存储效率之间寻求平衡也至关重要。

文献[48]提出了一种在多云间进行重复数据删除的新型架构 CloudShare，该架构用区块链管理多云，实现源端的重复数据删除。多个 CSP 在节省空间的利益驱动下进行合作，直接通过区块链交互，将文件元数据存入区块链以实现重复数据的高效精确删除，显著降低了每个云的存储成本，节省了用户的上传带宽。具体来说，每个 CSP 都有自己的用户并为其服务，但当 CSP1 的某用户上传了一个在 CSP2 中已存在的文件时，CSP1 会通过区块链发现，记录文件的所有权和文件的真实位置而不存储原始文件，并向 CSP2 支付一定费用。当该用户访问文件时，CSP1 向 CSP2 请问访问。

文献[49]也提出了一种基于区块链的源端重复数据删除方案，与 CloudShare 不同的是，该方案中重复数据删除在用

户本地通过智能合约执行，而 CloudShare 的重复数据删除过程对用户是透明的，由 CSP 执行而不使用智能合约。数据所有者在上传文件之前，先在本地计算文件标签，将其与从区块链获得的所有文件的标签进行对比，若存在重复的文件，则上传文件标签和原文件位置至智能合约并收到付费要求，数据所有者付费后将信息发给 CSP，CSP 将数据所有者付费记录发布在区块链上，待付费记录成功上链后将文件指针返回给数据所有者。若没有重复文件，则分块上传，创建待其他用户付费的智能合约，并收到每个文件块的指针。

文献[50]提出了一个基于区块链的安全重复数据删除方案 BDKM，该方案实现了可靠的 CK 管理，通过使用秘密共享方案将 CK 分成多块，并将其作为交易在区块链分发。下载 CK 时，只有有效的数据所有者才能从区块链恢复 CK，从区块链下载 CK 后，用户可以使用消息认证码验证 CK 的完整性，并最终恢复原始数据。但一定规模的区块链只能防止其存储的内容不被篡改，并不能防止内容泄露，如果区块链的部分节点泄露了一定数量的密钥块，则攻击者仍有可能从中获取完整密钥。

文献[51]针对 CE 算法的暴力攻击问题与密钥管理服务器的共谋攻击问题，提出了一个基于区块链的新型 CE 方案，使得 CE 密钥必须通过数据文件、用户私钥和系统随机数才能获得。用户私钥存储在区块链中，这增加了攻击者获取的难度，但密钥管理服务器仍然有泄露用户私钥的可能，为此，引入了任何一方都不可知的系统随机数，有效防止了共谋攻击。

为了保护重删后少量密文副本的完整性，防止重复伪造攻击（恶意用户篡改原始数据，但由于数据以密文形式存储，篡改行为难以被检测出）的发生，文献[52]用区块链记录原始数据信息，并从原始数据块延伸出一条侧链记录数据修改的所有信息，用于数据完整性破坏后追踪数据用户的真实身份。

还有研究者考虑了在执行重复数据删除的网络存储服务中，如何提高数据完整性验证的效率问题。在网络存储服务中，即使已经执行了重复数据删除，缺乏相互信任的数据所有者们仍会独立地为重复文件生成验证元数据。由于客户端重复数据删除和数据完整性验证过程都需要服务提供商存储数据块标签，因此两者的简单组合会导致标签数量快速增加，从而限制了系统的可扩展性，并违反了重复数据删除的目标。为了节省开销，一些已有方案在多个用户之间共享元数据，但都依赖不可信的第三方审计机构实现。此外，多个用户可能会重复检查他们共同持有数据的完整性，从而造成不必要的开销。文献[53]提出了一种基于区块链的重复数据删除前提下的完整性验证方案，使用区块链技术来消除不熟悉的用户之间的信任边界，避免了多个用户重复审计共同拥有的文件。该方案使用区块链作为用户、TPA 与存储服务提供方之间交互的工具，重复数据删除中的数据索引与数据完整性验证的中间数据均存储在区块链中，满足了多用户网络存储场景中的安全需求，同时减小了因数据重复导致的额外开销。

表 3 列出了从针对的问题与解决方案的角度，对上述基于区块链的云存储重复数据删除方案的总结。虽然通过区块链存储高冗余的元数据，可以进行可靠的重复数据删除以实现高效、低冗余的数据存储，这对于安全的云存储具有重大意义，但目前基于区块链的云存储重复数据删除的研究仍然很少，且现有研究大多是围绕用区块链解决重复数据删除过程中可能遇到的安全问题，并不是用区块链实现重复数据删除。

表 3　基于区块链的云存储重复数据删除方案对比

Table 3　Comparisons of blockchain-based cloud data deduplication

	问题	解决方法
CloudShare[48]	不可靠的重复数据检测	用区块链管理多云，重复数据删除对用户透明
文献[49]	不可靠的重复数据检测	用户在本地通过区块链进行重复数据删除
文献[50]	不可靠的密钥管理	通过秘密共享将密钥分块存储在区块链，当且仅当用户通过身份验证才可获得完整密钥
文献[51]	CE 暴力攻击和共谋攻击	通过引入用户私钥和系统随机数，增加获取 CE 密钥的难度
文献[52]	重复伪造攻击	用区块链记录原始数据信息，并从原始数据块延伸出一条侧链记录数据修改的所有信息
文献[53]	执行重复数据删除后完整性验证效率低下	使用区块链作为用户、TPA 与存储服务提供方之间交互的工具，用于存储中间数据，避免重复验证

3.4　基于区块链的云数据溯源

数据溯源指记录原始数据在整个生命周期内（从产生、传播到消亡）的演变信息和演变处理内容[54]。在云存储场景中，数据溯源能够记录云数据何时、何地以及用户如何在云存储服务器中存储、访问、修改和删除数据[55]，便于 CSP 进行数据管理，增强云数据的机密性、完整性和可用性。通过追溯与分析，数据溯源能够准确定位异常位置及检测云存储中的违规访问和恶意行为。

传统的云数据溯源是通过日志技术实现的[56]，然而其作用极其有限。云环境中，不同地理位置、不同组织中的多层软硬件可能会进行互操作，导致数据溯源需要从不同的来源收集信息，而日志是通过在给定的物理、虚拟或应用资源上执行软件生成的数据，信息有限。此外，传统云数据溯源方式依然存在中心化不可信的风险，且成本高、缺乏透明度。

防篡改、透明的特性使得区块链技术可以用于云环境中的数据溯源，由于区块链是分布式账本，多个云节点还可以自组织为区块链网络，维护一条溯源链以检测数据异常。

文献[57]提出了一种高效安全的云存储数据溯源方案 ESP，包含用户、CSP、审计员、认证服务器、区块链网络 5 个实体。认证服务器为每个用户分配一个密码，维护用户与密码

的列表，据此对用户进行身份验证。经过身份验证的用户每有一个文档处理请求，认证服务器就协助用户生成一条溯源记录并签名，向 CSP 证明该用户的合法身份，并将溯源记录集成在 Ethereum 公链中。用户向认证服务器支付 Ethereum 费用。ESP 引入了锁存窗口（WoL）的概念来评估方案的实用性。

文献[56,58]提出了一种基于区块链的数据溯源架构 ProvChain，该架构包括数据存储层、区块链层和溯源层。数据存储层包含一个 CSP，支持云存储应用；区块链层在 Tierion[59] 网络上实现，以文件作为数据单元存储每个数据文件的操作记录，可用于验证溯源数据库中数据的真实性；溯源层为本地扩展数据库，用于记录文件操作和查询以及恶意行为检测。溯源数据同时上传至区块链层和溯源层，在溯源数据通过区块链层的验证后，在溯源数据库中相应数据上更新验证状态，表明该数据的最新验证结果。每条数据记录与数据用户的 Hash ID 相关联，因此数据用户的真实身份不会被区块链网络节点和审计者所识别。但数据记录以明文形式存储在区块链中，仍然存在敏感数据泄露的风险。ProvChain 仅实现了一个 CSP 环境下的数据溯源，但可以扩展为多个 CSP 环境。ProvChain 中 CSP 根据用户的数据使用水平向区块链网络付费。然而只强调不可更改的时间戳和区块链收据的 ProvChain 对于数据保护来说是不够的。文献[60]提供了 Ethereum 区块链网络下的区块结构的设计和区块生成的过程，该设计增强了溯源系统中溯源日志的可见性和可追溯性。

在多云环境中，云存储应用将需要解决互操作、跨提供商数据共享和管理问题，为了增强安全性，仍需从不同 CSP 和不同云存储应用中收集元数据并进行数据溯源。多云环境中节点之间的数据交换/共享是常见的场景，因为一项服务通常由多个服务提供商支持。物理存储的变化可能导致更多的数据泄露漏洞，例如恶意数据挖掘、网络监控、干扰或欺骗攻击。实现可靠的数据溯源是提供安全云服务的一个重要问题。然而，由于数据可能会被传输到不同的供应商，而不是在一个空间中进行存储和操作，因此数据所有者很难识别服务背后的数据流。

ProvChain 的研究团队在文献[61-62]中提出了在联合云环境下与 ProvChain 类似的数据溯源架构 BlockCloud，该架构包含区块链层、溯源层和联合云环境层。BlockCloud 中，区块链网络由云用户自组织形成。与 ProvChain 不同的是，由于性能、空间和隐私的限制，详细的溯源数据不能直接存储在区块链层，只能存储在溯源数据库中，区块链层只存储摘要信息，这进一步增强了元数据的机密性。在 BlockCloud 中使用权益证明（PoS）共识机制，以减轻传统工作证明（PoW）共识机制所需的计算需求开销。BlockCloud 的溯源层与溯源审计器和联合云服务交互，前者用于插入/查询起源数据以及检测恶意行为，后者用于区块验证。

文献[63]更有针对性地提出了使用区块链来管理云数据中心之间的数据传输方案 BDP2。BDP2 将每个云数据中心视为区块链网络中的一个节点，将云之间的数据传输行为表述为一个区块链交易。该方案主要讨论两种数据传输类型，分别为数据迁移和数据复制，并设计了两种核心算法，一种是任务类型检测算法，用于判断数据的传输类型，另一种是验证算法，用于验证传输有效性。由于只用到了区块链的数据追踪功能，BDP2 在 Ethereum 私有链上实现。

由于物联网设备数据存储和处理能力有限，工业控制、智能电网、环境监测系统等应用高度依赖以云为中心的物联网网络，这样的网络系统需要高度可信的数据来保证准确和及时的决策。然而，由于系统的复杂性，数据的可靠性难以保证。文献[64]用 Hyperledger Fabric 构建私有链，为以云为中心的物联网网络提出了一个安全的数据溯源框架，在该框架中，设备元数据的 hash 值存储在区块链，而实际数据存储在链外，即云中，这使得该框架具有高度可扩展性。多个智能合约自动执行以保证链上数据的有效性。该框架使用简化拜占庭容错（SBFT）共识算法。拜占庭容错（PBFT）算法对中央处理器友好，能够按照物联网网络的要求每秒处理大量事务[65]。文献[66]则在 Hyperledger Fabric 中，通过将物联网节点的部分工作转移到边缘节点上，实现了以轻量级的安全开销支持区块链网络上的高吞吐量事务。

通过数据溯源进行云取证是发现网络犯罪的有效手段，但集中化的证据收集和保存不能保证数据证据的高可靠性。为解决这一问题，文献[67]提出了一种新的数据取证体系结构，该体系结构结合了软件定义网络（SDN）和面向 LaaS 云的区块链技术，在 SDN 控制器中为每个数据创建区块，将原始数据的演变历史作为元数据存储在区块中，最后通过构建从区块链收集的证据逻辑图来进行证据分析；为了保护系统免受未授权用户的攻击，其还提出了基于安全环验证的身份认证方案，与 ESP[49] 的密码方案相比，安全性大大增强。该方案在由 Java 创建的基于 PoW 的区块链环境中实现，为了提高效率，每个块中的 MHT 用 SHA-3 算法生成。由于云存储的分布式特性，收集电子证据可能需要多方合作，例如辖区 A 的调查员 Alice 需要收集的目标证据由辖区 B 的管理员 Bob 维护，则 Alice 需要向 Bob 请求数据，Bob 收集证据后，电子证据可能由第三个人 Eve 提交给 Alice，在这种情况下，证据的传输过程记录十分重要。文献[68]针对上述问题，利用区块链技术和群签名技术，将电子证据请求和提交记录用 Chainpoint 协议存储在类似 Bitcoin 或 Ethereum 的区块链网络中，隐藏发送者和接受者的身份，增强了云取证的可信度，但该方案仍存在中心化审计机构和证书颁发机构的限制。

文献[69]提出了分布式云存储数字资产安全传输架构 DistProv，使用 IPFS 在发送方和接受方之间存储和传输数字资产。其中，两个不可信方都拥有一个密钥对和区块链地址以进行身份验证和与区块链进行交互。两方交换由加密算法保护的敏感文档，并使用 Ethereum 智能合约通过零知识证明验证访问权限，同时将数字资产的溯源数据作为交易发布在区块链。DistProv 建立在私有链上，使用权威证明（PoA）共识算法。

表 4 列出了从区块链交易内容、用户身份隐私机制、云环

境、共识算法、区块链平台和类型方面对上述基于区块链的云数据溯源方案进行对比分析的结果。区块链的可信与智能合约的自动化执行特点使其适用于单云、多云或多数据中心环境识算法和用性能较高的区块链平台来解决。此外，由于数据溯源需收集数据在整个生命周期内的演变过程，用户身份的保护尤为重要，否则泄露身份与数据的对应关系会对隐私造成极大威胁。区块链本身的匿名性仍不能防止恶意人员从所有公开的交易中分析出用户的个人信息，因此最好结合其他密码学方案保证用户身份的隐私性，如环签名和群签名等。

表4　基于区块链的云数据溯源方案对比

Table 4　Comparisons of blockchain-based cloud data provenance

	区块链交易内容	用户身份	云环境	共识算法	区块链平台	区块链类型
ESP[57]	文档修改记录	密码	单云存储	PoW	Ethereum	公有链
ProvChain[56,58]	数据文件操作记录	hash ID	单云存储（可扩展为多云）	N/A	Tierion	公有链
文献[60]	数据文件操作记录	hash ID	单云存储	PoW	Ethereum	私有链
BlockCloud[61-62]	数据文件操作记录	hash ID	多云存储	PoS	Tierion	公有链
BDP2[63]	数据传输信息	N/A	多数据中心	PoW	Ethereum	私有链
文献[64]	设备元数据散列值	N/A	以云为中心的物联网	SBFT	Hyperledger Fabric	私有链
BlockTrack-L[66]	设备元数据散列值	N/A	以云为中心的物联网	PBFT	Hyperledger Fabric	私有链
文献[67]	证据元数据(SHA-3)	安全环验证	使用SDN实现云取证的LaaS云环境	PoW	Java创建	私有链
文献[68]	电子证据传输记录	群签名	单云存储	N/A	Tierion	公有链
DistProv[69]	数字资产	密钥	P2P云存储	PoA	Ethereum	私有链

4　技术挑战分析

将区块链技术运用于云存储安全能够增强系统的安全性，但是已有的研究以及未来的研究中可能会出现以下几种挑战。

4.1　链上数据来源不可信

尽管区块链能够有效保证链上数据的真实可信，但数据在上传到区块链之前有可能是已经被篡改了的数据。当前基于区块链的云存储安全方案均基于上传数据至区块链的实体是可信的假设，但实际情况可能并没有这么理想。对于此类情形，应对提供数据的实体的可靠性进行分析论证。

4.2　云端需开发接口

在当前基于区块链的云存储安全解决方案中，云端除了提供基础存取数据服务外，还需主动提供额外的计算和数据接口来与用户、第三方机构或者区块链进行交互，但大型的商业化CSP几乎不可能因为其他机构的服务需求开发接口[31]。一种解决思路是，只使用CSP的存取数据服务，所有的接口设置在客户端或代理上。

4.3　安全隐私挑战

区块链技术本身在安全隐私方面存在诸多挑战，将区块链应用于云存储安全机制中会带来一定效益，但也会带来其他方面的安全隐私问题[70-71]。

(1)用户身份可能暴露。以比特币为例，虽然区块链中用户用公钥生成hash，再用公钥hash生成地址作为身份标识以实现匿名性，即使每次交易都用一次性地址，但多个地址连接到同一用户，恶意人员依然能够从与同一个用户相关联的所有交易中分析出对应用户的行为特性、交易伙伴或其他敏感信息。解决该问题的方法有混币服务[72](mixing)、零知识证明[73]、环签名技术[74]等。

(2)链上数据隐私容易泄露。区块链是公开账本，所有接入区块链的节点都能看到所有交易数据。数据在上传到区块链之前应被加密。在对整体数据进行统计分析时，可采用差别隐私[75]方法。

(3)用户失去对数据的控制权。区块链只能增加不能删减，这意味着将用户的个人数据上传到区块链后无法删除。

4.4　吞吐量挑战

目前部分区块链平台确认一条交易的时间较长，吞吐量不高。由于云数据量庞大，在部分云存储场景中，运用区块链会制约整个系统的效率。通过设计新的共识算法缩短共识时间，或者设计新型区块链结构(如侧链[76])能够改善此问题。

4.5　个人用户可能缺少硬件条件

区块链操作对计算机内存、电池、处理器等有较高要求。本文提到的部分基于区块链的云存储安全方案在云用户之间构造区块链网络，或者需要云用户与区块链进行交互，但部分云用户是个人，并不具备这样操作的硬件条件。避免用户与区块链的交互，将区块链操作转移到一个代理上，可以避免此问题。

5　总结与展望

5.1　总结

本文首先对云存储安全和区块链的安全性进行了概述，云存储安全需要依赖某些安全机制来保障，但传统安全机制存在中心化、数据不可信以及场景中特有的问题，带有去中心化、持久性、匿名性、可审计性特点的区块链技术可以解决以上威胁。因此，对云存储中研究得较多的基于区块链的访问控制、完整性验证、重复数据删除以及数据溯源技术进行了总结，图6给出了对各个方面运用区块链来解决的问题的总结。最后，对基于区块链的云存储安全进行了技术挑战分析。云存储安全对于个人、企业和组织的数据至关重要，区块链技术能够弥补传统云访问控制、完整性验证、重复数据删除、数据溯源等云存储安全机制的固有缺点，在云存储安全中有广泛的应用前景。在将区块链技术与云存储安全结合的过程中，

注意解决运用区块链带来的各类挑战，能够发挥区块链技术的最大应用价值。

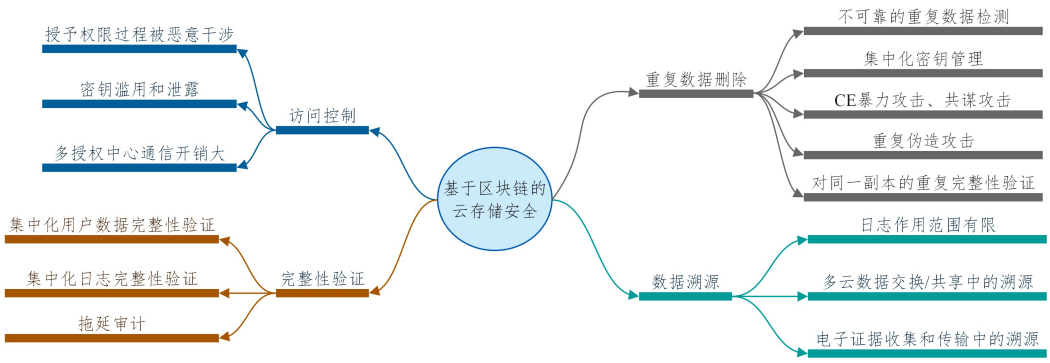

图 6 基于区块链的云存储安全总结

Fig. 6 Summary of blockchain-based cloud storage security mechanism

5.2 展望

下面对区块链与云存储的结合在未来可能的研究方向进行展望。

(1)基于差别隐私的可信云异常行为检测

在基于区块链的云存储数据溯源中，通过对历史数据进行统计分析，可以识别出与整体不符的异常数据，从而检测出异常状态。该方案要求存储在区块链上的历史数据能够揭露信息，否则无法从中分析出趋势和异常。但同时，用户的个人隐私也面临泄露的风险。差别隐私技术在数据中加入随机噪声，使得个人数据不会被准确泄露，但整体统计结果不会受到影响。将差别隐私技术用于区块链云异常行为检测中，可在保证历史数据可信的同时，增强云用户个人信息的安全性。

(2)基于区块链的云代理联盟

为了能够合理地协调云存储资源，单代理多云架构通过代理将用户和各大云存储提供商联合起来，用户通过代理实现成本和服务水平的最优化。然而，单个代理由于性能瓶颈，管理的云数量有限。如果将多个代理联合起来，扩大云存储的资源范围，通过在云代理联盟管辖的总体范围内进行重复数据删除、数据布局优化，可以获得超过单个代理所能达到的服务水平，这就形成了多代理多云架构。多代理多云架构须在代理间实现数据共享，但其中存在数据一致性保障的问题。分布式架构的区块链技术能够成为实现云代理联盟的途径。云代理联盟可形成一个区块链网络，每个云代理可作为一个节点自主加入到区块链网络中，利用区块链提供的可靠全局视图进行重复数据删除等工作，节省存储资源，增加用户数量。

(3)云存储中的跨链融合

跨链融合可能会成为区块链与云存储结合的未来趋势，其原因有两点：1)面对云中庞大的用户量和数据量，一条不断增长的区块链很难长远地满足存储量的需求；2)在多个功能之间实现数据共享可以减少不必要的存储与计算冗余。但区块链网络之间相互独立，存在“信息孤岛”，实现区块链网络间的互通互联与协作是一项挑战。

参 考 文 献

[1] CHAI Q, GONG G. Verifiable symmetric searchable encryption for semi-honest-but-curious cloud servers[C]// 2012 IEEE International Conference on Communications (ICC). IEEE, 2012: 917-922.

[2] ALMORSY M, GRUNDY J, MÜLLER I. An analysis of the cloud computing security problem[J]. arXiv: 1609.01107, 2016.

[3] WU J, PING L, GE X, et al. Cloud storage as the infrastructure of cloud computing[C]// 2010 International Conference on Intelligent Computing and Cognitive Informatics. IEEE, 2010: 380-383.

[4] FU Y, LUO S, SHU J. Survey of Secure Cloud Storage System and Key Technologies[J]. Journal of Computer Research and Development, 2013(1): 136-145.

[5] NAKAMOTO S. Bitcoin: A peer-to-peer electronic cash system [J/OL]. Decentralized Business Review, 2008: 21260. https://www.researchgate.net/publication/228640975_Bitcoin_A_Peer-to-Peer_Electronic_Cash_System.

[6] ZHENG Z, XIE S, DAI H, et al. An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends[C]// IEEE International Congress on Big Data. Piscaway: IEEE, 2017.

[7] SEGURA S D, PÉREZ-SOLÀ C, NAVARRO-ARRIBAS G, et al. Analysis of the Bitcoin UTXO Set[C]// 22nd International Conference on Financial Cryptography and Data Security (FC 2018). 2018.

[8] Protocol Labs. Filecoin: A Decentralized Storage Network[OL]. https://filecoin.io/filecoin.pdf.

[9] BENET J. Ipfs-content addressed, versioned, p2p file system [J]. arXiv: 1407.3561, 2014.

[10] WILKINSON S. Storj A Peer-to-Peer Cloud Storage Network [OL]. http://storj.io/storj.pdf.

[11] VORICK D, CHAMPINE L. Sia: Simple decentralized storage [OL]. https://blockchainlab.com/pdf/whitepaper3.pdf.

[12] LAMBDA P. A Blockchain Infrastructure Providing Unlimited Storage Capabilities[OL]. https://www.lambdastorage.com/doc/Lambda%E7%BB%8F%E6%B5%8E%E7%99%BD%E7%9A%AE%E4%B9%A6.pdf.

[13] TAVIZI T, SHAJARI M, DODANGEH P. A usage control based architecture for cloud environments[C]// 2012 IEEE 26th International Parallel and Distributed Processing Symposium

Workshops & PhD Forum. IEEE,2012:1534-1539.

[14] LIN G Y,HE S,HUANG H,et al. Access control security model based on behavior in cloud computing environment[J]. Journal on Communications,2012,33(3):59-66.

[15] BETHENCOURT J,SAHAI A,WATERS B. Ciphertext-policy attribute-based encryption[C]// 2007 IEEE Symposium on Security and Privacy (SP'07). IEEE,2007:321-334.

[16] SOHRABI N,YI X,TARI Z,et al. BACC:blockchain-based access control for cloud data[C]// Proceedings of the Australasian Computer Science Week Multiconference. 2020:1-10.

[17] GUO J,YANG W,LAM K Y,et al. Using blockchain to control access to cloud data[C]// International Conference on Information Security and Cryptology. Springer,Cham,2018:274-288.

[18] YANG C,TAN L,SHI N,et al. AuthPrivacyChain:A blockchain-based access control framework with privacy protection in cloud[J]. IEEE Access,2020,8:70604-70615.

[19] SUKHODOLSKIY I,ZAPECHNIKOV S. A blockchain-based access control system for cloud storage[C]// 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus). IEEE,2018:1575-1578.

[20] WANG S,WANG X,ZHANG Y. A secure cloud storage framework with access control based on blockchain[J]. IEEE Access,2019,7:112713-112725.

[21] QIN X,HUANG Y,YANG Z,et al. A Blockchain-based access control scheme with multiple attribute authorities for secure cloud data sharing[J]. Journal of Systems Architecture,2021,112:101854.

[22] GAO S,PIAO G,ZHU J,et al. TrustAccess:A trustworthy secure ciphertext-policy and attribute hiding access control scheme based on blockchain[J]. IEEE Transactions on Vehicular Technology,2020,69(6):5784-5798.

[23] GUO L,YANG X,YAU W C. TABE-DAC:Efficient Traceable Attribute-Based Encryption Scheme With Dynamic Access Control Based on Blockchain[J]. IEEE Access,2021,9:8479-8490.

[24] QIN X,HUANG Y,YANG Z,et al. LBAC:A lightweight blockchain-based access control scheme for the internet of things [J]. Information Sciences,2021,554:222-235.

[25] KUMAR R,PALANISAMY B,SURAL S. BEAAS:Blockchain Enabled Attribute-Based Access Control as a Service[C]// 2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC). IEEE,2021:1-3.

[26] WANG S,ZHANG Y,ZHANG Y. A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems [J]. IEEE Access,2018,6:38437-38450.

[27] SCHIAVO F P,SASSONE V,NICOLETTI L,et al. Faas:Federation-as-a-service[J]. arXiv:1612.03937,2016.

[28] FERDOUS M S,MARGHERI A,PACI F,et al. Decentralised runtime monitoring for access control systems in cloud federations[C]// 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). IEEE,2017:2632-2633.

[29] ALANSARI S,PACI F,SASSONE V. A distributed access control system for cloud federations[C]// 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). IEEE,2017:2131-2136.

[30] PRIYADHARSHINI B,PARVATHI P. Data integrity in cloud storage[C]// IEEE-international conference on advances in engineering,science and management (ICAESM-2012). IEEE,2012:261-265.

[31] YANG C. Research on Blockchain-based Cloud Storage Data Integrity Detection[D]. School of Computer Science and Engineering,2020.

[32] GAETANI E,ANIELLO L,BALDONI R,et al. Blockchain-based database to ensure data integrity in cloud computing environments[C]// the First Italian Conference on Cybersecurity (ITASEC17). 2017:146-155.

[33] ZIKRATOV I,KUZMIN A,AKIMENKO V,et al. Ensuring data integrity using blockchain technology[C]// 2017 20th Conference of Open Innovations Association (FRUCT). IEEE,2017:534-539.

[34] WEI P C,WANG D,ZHAO Y,et al. Blockchain data-based cloud data integrity protection mechanism[J]. Future Generation Computer Systems,2020,102:902-911.

[35] XUE J,XU C,ZHAO J,et al. Identity-based public auditing for cloud storage systems against malicious auditors via blockchain [J]. Science China Information Sciences,2019,62(3):32104.

[36] ZHANG G,YANG Z,XIE H,et al. A secure authorized deduplication scheme for cloud data based on blockchain[J]. Information Processing & Management,2021,58(3):102510.

[37] LI S,LIU J,YANG G,et al. A Blockchain-Based Public Auditing Scheme for Cloud Storage Environment without Trusted Auditors[J]. Wireless Communications and Mobile Computing,2020,2020:8841711.

[38] LI J,WU J,JIANG G,et al. Blockchain-based public auditing for big data in cloud storage[J]. Information Processing & Management,2020,57(6):102382.

[39] PINHEIRO A,CANEDO E D,DE SOUSA R T,et al. Monitoring File Integrity Using Blockchain and Smart Contracts[J]. IEEE Access,2020,8:198548-198579.

[40] ZHANG C,XU Y,HU Y,et al. A blockchain-based multi-cloud storage data auditing scheme to locate faults[J]. IEEE Transactions on Cloud Computing,2021:3057771.

[41] YUE D,LI R,ZHANG Y,et al. Blockchain based data integrity verification in P2P cloud storage[C]// 2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS). IEEE,2018:561-568.

[42] XUE J,XU C,ZHANG Y,et al. DStore:a distributed cloud storage system based on smart contracts and blockchain[C]// International Conference on Algorithms and Architectures for Parallel Processing. Cham:Springer,2018:385-401.

[43] WANG J,PENG F,TIAN H,et al. Public auditing of log integrity for cloud storage systems via blockchain[C]// International Conference on Security and Privacy in New Computing Environments. Cham:Springer,2019:378-387.

[44] ZHANG Y, XU C, LIN X, et al. Blockchain-based public integrity verification for cloud storage against procrastinating auditors [J]. IEEE Transactions on Cloud Computing, 2019: 2908400.

[45] YUAN H, CHEN X, WANG J, et al. Blockchain-based public auditing and secure deduplication with fair arbitration[J]. Information Sciences, 2020, 541: 409-425.

[46] DOUCEUR J R, ADYA A, BOLOSKY W J, et al. Reclaiming space from duplicate files in a serverless distributed file system [C]// Proceedings 22nd International Conference on Distributed Computing Systems. IEEE, 2002: 617-624.

[47] BELLARE M, KEELVEEDHI S, RISTENPART T. Message-locked encryption and secure deduplication[C]// Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer, Berlin, Heidelberg, 2013: 296-312.

[48] LI Y, ZHU L, SHEN M, et al. Cloudshare: towards a cost-efficient and privacy-preserving alliance cloud using permissioned blockchains[C]// International Conference on Mobile Networks and Management. Springer, Cham, 2017: 339-352.

[49] LI J, WU J, CHEN L, et al. Deduplication with blockchain for secure cloud storage[C]// CCF Conference on Big Data. Springer, Singapore, 2018: 558-570.

[50] ZHANG G, XIE H, YANG Z, et al. BDKM: A Blockchain-Based Secure Deduplication Scheme with Reliable Key Management [J]. Neural Processing Letters, 2021(3): 1-18.

[51] ZHANG G, YANG Z, XIE H, et al. A secure authorized deduplication scheme for cloud data based on blockchain[J]. Information Processing & Management, 2021, 58(3): 102510.

[52] HUANG H, CHEN Q, ZHOU Y, et al. Blockchain-Based Secure Cloud Data Deduplication with Traceability[C]// International Conference on Blockchain and Trustworthy Systems. Springer, Singapore, 2020: 295-302.

[53] XU Y, ZHANG C, WANG G, et al. A blockchain-enabled deduplicatable data auditing mechanism for network storage services [J/OL]. IEEE Transactions on Emerging Topics in Computing, 2020. https://www.researchgate.net/publication/342539890_A_Blockchain-enabled_Deduplicatable_Data_Auditing_Mechanism_for_Network_Storage_Services.

[54] MING H, ZHANG Y, FU X. Survey of Data Provenance[J]. Journal of Chinese Computer Systems, 2012(9): 1917-1923.

[55] GAI K, GUO J, ZHU L, et al. Blockchain meets cloud computing: a survey[J]. IEEE Communications Surveys & Tutorials, 2020, 22(3): 2009-2030.

[56] LIANG X, SHETTY S S, TOSH D, et al. ProvChain: Blockchain-based Cloud Data Provenance[M]. Blockchain for Distributed Systems Security, 2019: 67-94.

[57] ZHANG Y, LIN X, XU C. Blockchain-based secure data provenance for cloud storage[C]// International Conference on Information and Communications Security. Springer, Cham, 2018: 3-19.

[58] LIANG X, SHETTY S, TOSH D, et al. Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability[C]// 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID). IEEE, 2017: 468-477.

[59] Tierion: Blockchain Proof Engine | API[OL]. 2018. https://tierion.com.

[60] SIFAH E B, XIA Q, AGYEKUM K O B O, et al. A Blockchain Approach to Ensuring Provenance to Outsourced Cloud Data in a Sharing Ecosystem [J/OL]. IEEE Systems Journal, 2021: 3068224. https://ieeexplore.ieee.org/document/9405789.

[61] SHETTY S, RED V, KAMHOUA C, et al. Data provenance assurance in the cloud using blockchain[C]// Disruptive Technologies in Sensors and Sensor Systems. International Society for Optics and Photonics, 2017, 10206: 1020601.

[62] TOSH D, SHETTY S, LIANG X, et al. Data provenance in the cloud: A blockchain-based approach[J]. IEEE Consumer Electronics Magazine, 2019, 8(4): 38-44.

[63] LI H, GAI K, FANG Z, et al. Blockchain-enabled data provenance in cloud datacenter reengineering[C]// Proceedings of the 2019 ACM International Symposium on Blockchain and Secure Critical Infrastructure. 2019: 47-55.

[64] ALI S, WANG G, BHUIYAN M Z A, et al. Secure data provenance in cloud-centric internet of things via blockchain smart contracts[C]// 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/ SCI). IEEE, 2018: 991-998.

[65] SHAFAGH H, BURKHALTER L, HITHNAWI A, et al. Towards blockchain-based auditable storage and sharing of IoT data[C]// Proceedings of the 2017 on Cloud Computing Security Workshop. 2017: 45-50.

[66] SIDDIQUI M S, ALI T, NADEEM A, et al. BlockTrack-L: A lightweight blockchain-based provenance message tracking in IoT[J]. International Journal of Advanced Computer Science and Applications, 2020, 11(4): 463-470.

[67] POURVAHAB M, EKBATANIFARD G. Digital forensics architecture for evidence collection and provenance preservation in iaas cloud environment using sdn and blockchain technology[J]. IEEE Access, 2019, 7: 153349-153364.

[68] ZHANG Y, WU S, JIN B, et al. A blockchain-based process provenance for cloud forensics[C]// 2017 3rd IEEE International Conference on Computer and Communications (ICCC). IEEE, 2017: 2470-2473.

[69] GOURU N, VADLAMANI N L. DistProv-Data Provenance in Distributed Cloud for Secure Transfer of Digital Assets with Ethereum Blockchain using ZKP[M]// Cyber Warfare and Terrorism: Concepts, Methodologies, Tools, and Applications. IGI Global, 2020: 866-890.

[70] BERNABE J B, CANOVAS J L, HERNANDEZ-RAMOS J L, et al. Privacy-preserving solutions for blockchain: Review and challenges[J]. IEEE Access, 2019, 7: 164908-164940.

[71] JOSHI A P, HAN M, WANG Y. A survey on security and pri-

vacy issues of blockchain technology[J]. Mathematical Foundations of Computing,2018,1(2):121-147.

[72] CHAUM D L. Untraceable electronic mail,return addresses,and digital pseudonyms[J]. Communications of the ACM, 1981, 24(2):84-90.

[73] GOLDREICH O,MICALI S,WIGDERSON A. Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems[J]. Journal of the ACM (JACM), 1991,38(3):690-728.

[74] RIVEST R L,SHAMIR A,TAUMAN Y. How to leak a secret: Theory and applications of ring signatures[M]// Theoretical Computer Science. Springer,Berlin,Heidelberg,2006:164-186.

[75] DWORK C,ROTH A. The algorithmic foundations of differential privacy[J]. Foundations and Trends in Theoretical Computer Science,2014,9(3/4):211-407.

[76] BACK A,CORALLO M,DASHJR L,et al. Enabling blockchain innovations with pegged sidechains[OL]. http://www.opensciencereview.com/papers/123/enablingblockchain-innovations-with-pegged-sidechains,2014,72.

XU Kun, born in 1997, postgraduate. Her main research interests include cloud storage and blockchain.

CHEN Wei-wei, born in 1967, professor, is a member of China Computer Federation . Her main research interests include services computing and cloud computing.