软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, [doi: 10.13328/j.cnki.jos.005771]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于区块链的大数据访问控制机制*

刘敖迪[1,2], 杜学绘[1,2], 王 娜[1,2], 李少卓[1,2]

[1](信息工程大学,河南 郑州 450001)

[2](河南省信息安全重点实验室,河南 郑州 450001)

通讯作者: 杜学绘, E-mail: dxh37139@sina.com

摘 要: 针对大数据资源来源广泛、动态性强且呈现出分布式管理的特点,当前主流集中式访问控制机制存在权限管理效率低、灵活性不足、扩展性差等不足.基于此,本文以ABAC模型为基础,提出了一种基于区块链的大数据访问控制机制.首先,对区块链技术的基本原理进行描述并对基于属性的访问控制模型进行形式化的定义;然后,提出了一个基于区块链技术的大数据访问控制架构,并对访问控制的基本框架与流程进行了详细的阐述与分析;同时,对基于区块链事务的访问控制策略及实体属性信息管理方法进行了说明,以此保证访问控制信息的不可篡改性、可审计性和可验证性;随后,采用基于智能合约的访问控制方法实现对大数据资源由用户驱动、全程透明、动态、自动化的访问控制;最后,通过仿真实验验证了该机制的有效性,并对本文的研究内容进行总结与展望.

关键词: 大数据安全;访问控制;区块链;ABAC 模型;智能合约

中图法分类号: TP311

A Blockchain-Based Access Control Mechanism for Big Data

LIU Ao-Di[1,2], DU Xue-Hui[1,2], WANG Na[1,2], LI Shao-Zhuo[1,2]

[1](Information Engineering University, Zhengzhou 450001, China)

[2](He'nan Province Key Laboratory of Information Security, Zhengzhou 450001, China)

Abstract: In the terms of the wide source, large dynamics, and distributed management characteristics of big data resources, the current mainstream centralized access control mechanisms have some shortcomings, such as low efficiency, insufficient flexibility, and poor scalability. Therefore, this paper proposes a blockchain-based big data access control mechanism based on the ABAC model. First, the fundamental principle of blockchain technology is described and the attribute-based access control model is formalized; Then, we presents a big data access control architecture based on blockchain technology, and analyzes the basic framework and flow of access control. At the same time, to ensure the access control information is tamper-resistant, auditability and verifiability, the paper also describes the transaction-based access control policy and entity attribute information management methods in detail. In addition, a smart contract-based access control method is used to implement user-driven, transparent, dynamic, and automated access control for big data resources. Finally, simulation experiments validate the effectiveness of this mechanism, and then we summarize and prospect the views presented in this paper.

Key words: big data security; access control; blockchain; ABAC model; smart contract

目前大数据已经在生产生活中得到了非常广泛的应用,大数据所带来的社会变革也已深入到我们生活的方方面面.例如,通过分析大量病人的临床"医疗大数据[1]",能够使医生更好的理解病症,实现病情的精准诊断与

* 基金项目:国家重点研发计划(2018YFB0803603,2016YFB0501901);国家自然科学基金(61802436); 河南省自然科学基金(162300410334)

Foundation item: National Key Research and Development Program of China(2018YFB0803603,2016YFB0501901); National Natural Science Foundation of China (61802436); Natural Science Foundation of Henan Province of China(162300410334)

收稿时间: 2018-06-09; 修改时间: 2018-08-28; 采用时间: 2018-12-14; jos 在线出版时间: 2019-04-10

CNKI 网络优先出版: 2019-04-09 17:32:15, http://kns.cnki.net/kcms/detail/11.2560.TP.20190409.1731.001.html

治疗,并且能够提高流行病的预警能力,有助于采取有效措施预防流行病的爆发;通过将“能源大数据[2]”与人口、地理、气象等相关领域数据进行整合及分析利用,能够实现能源生产、运营、消费产出的最大化,促进能源产业发展及商业模式创新;通过对大量网络交易及行为等“金融大数据[3]”的分析,可提高金融企业在资本管理、交易执行、安全和反欺诈等方面的数据洞察力,提高企业核心竞争力.由此可见,在大数据时代,数据已成为一种能够流动的资产宝库,通过分析、利用大数据能够创造出巨大的社会和经济价值,并且数据量越大、来源越广泛,产生的价值也会越大.

然而,大数据在带来新的发展机遇的同时,也面临着严峻的数据安全挑战.一方面,为了更好的挖掘数据中的价值,离不开分布式数据源间数据的流通与共享,但由此必然会打破原有数据管理的安全边界,增加了数据在共享过程中所面临的安全风险;另一方面,由于大数据资源具有巨大的经济价值,针对大数据资源的窃取、攻击与滥用等行为越来越严重,对国家及相关机构数据安全防护能力提出了更高的要求.在多因素压力下,导致大数据安全事故频发.特别是近期引起广泛关注的 Facebook 数据外泄事件,使大数据资源的非授权使用问题引起了安全专家的广泛关注.2018 年 3 月曝出,Facebook 超过 5000 万用户信息数据遭到外泄,一家名为 Cambridge Analytica 的数据分析公司在未经过用户授权的前提下搜集了超过 5000 万用户的个人数据,并利用该数据建立数学模型来分析用户的政治偏好,通过有针对性的向美国选民投放精准政治广告的形式,从而影响了 2016 年美国总统大选的结果.由此可以看出,大数据的非授权共享不单会影响用户自身的数据安全,更会对国家安全造成严重的安全威胁,实现安全、可控的大数据资源流通与共享是大数据应用及其发展所面临的核心科学问题.

作为保护数据安全的重要手段,访问控制技术通过对用户权限进行管理,使合法用户只能依照其所拥有的权限访问系统内相应数据,禁止用户对数据的非授权访问,从而保障数据安全和业务系统的正常运转.在大数据时代,访问控制技术[4, 5]仍将作为保护大数据安全、可控共享的重要手段.但是,大数据的分布式、动态性环境[6]使大数据的管理场景和安全需求变得更加复杂,传统的访问控制技术已不再适用.大数据访问控制面临访问控制策略制定以及授权管理难度增加、访问控制客体描述困难、受访问数据的主体集合构成复杂、访问控制对数据客体中个人隐私保护难度高、缺乏对大数据分析过程安全性考虑等若干挑战[7].不同于针对上述挑战的研究,本文主要针对其分布式访问控制需求、访问控制动态性需求这 2 个易被忽略的需求挑战展开研究.

(1)分布式访问控制需求挑战,大数据由众多分布式的数据源汇聚生成,不同的数据源可能位于不同的机构、公司以及组织内部,不同机构基于自身的数据安全保护需求,不可能直接与其它机构共享所有大数据资源,大数据资源具有藩篱化特征,需要在分布式复杂环境下实现对共享大数据资源有效的访问控制,其分布式体现在如下两个方面.

a)访问控制策略分布式制定:为提高访问控制策略的管理效率,传统集中式的访问控制机制需要由安全管理员统一进行策略管理,难以满足分布式大数据的安全共享需求,大数据共享与流通需要由资源的拥有者制定并维护其所拥有数据资源的访问控制策略,实现由拥有者驱动的访问控制策略分布式制定.另外,由于开放共享更有助于大数据价值的挖掘,这就要求大数据资源的策略需要进行可信公开,便于资源使用方进行策略的查询与其真实性验证,促进大数据资源在拥有访问权限的资源使用方间高效流通和共享.

b)访问控制分布式判决:大数据不宜只存在一个集中式的权限判决点,集中式权限判决不符合大数据的应用场景,大数据的流通与共享是一个多方参与、群智感知的应用过程,需要参与方的多元交互,才能充分挖掘大数据价值,提高大数据资源利用率.并且,传统由第三方机构进行的单一权限判决可能存在用户不可知的越权行为,存在权限判决透明度的问题.另外,当单一权限判决点发生故障时,将导致整个大数据系统停止运转,也存在单点故障的问题.

(2)访问控制动态性需求挑战,大数据动态产生且增长速度快,需要对动态生成的新数据及时进行访问控制管理.传统集中式的访问控制机制一般是针对静态资源进行管理,动态扩展能力弱,灵活性低且存在数据访问控制管理的滞后性,无法对新生成的数据资源进行高时效的访问控制管理.

近年来,区块链[8]作为一项从数字加密货币领域诞生的新兴技术引起了各领域研究人员的广泛关注.传统社会的信任建立在可信第三方信用“背书”的信任机制下,而区块链技术通过将 P2P 网络、密码学技术、共识机

制以及智能合约等多种技术进行深度整合解决了去中心化系统节点间信任建立的问题,实现了去中心化、分布式、信息不可篡改的信任建立机制,能够在信息传输的同时完成价值的转移.区块链技术的分布式架构与智能合约技术恰好与大数据环境下分布式、动态访问控制需求相吻合,基于此,本文针对大数据访问控制中所面临的挑战,基于区块链的事务管理和智能合约技术实现了对分布式环境下大数据资源动态、灵活的访问控制.本文的主要工作包括:将区块链技术与 ABAC 模型[9, 10]相结合,提出了一种基于区块链的分布式大数据访问控制机制 BBAC-BD(Blockchain-based Access Control Mechanism for Big Data Environment),通过改进区块链事务存储结构,利用区块链事务来实现访问控制策略的分布式管理,针对策略管理效率较低的问题,提出了基于 Bloom Filter 的策略管理方法,以实现访问控制策略的快速检索;利用智能合约技术实现策略的分布式自动、可信判决,以此实现用户对大数据资源的灵活管控,在禁止非法用户对数据资源访问的同时,提高大数据资源的共享与流通效率,实现面向分布式大数据资源的高效、透明、安全的自动化访问控制.

本文将区块链应用于访问控制技术主要有以下 5 个方面的优势:

(1) 资源的管理使用权真正掌握在资源拥有者手中,存储在区块链上的策略信息对所有主体可见,策略的可信公开更利于促进大数据资源的共享,有助于大数据资源价值的挖掘与利用.

(2) 基于智能合约能够实现对大数据资源自动化、可信的访问控制,无需安全管理员人为参与,基于资源拥有者发布的策略进行访问控制,判决过程公开透明.

(3) 基于 ABAC 模型,策略由资源拥有者发布到区块链上,随着大数据资源的动态变化,资源拥有者可及时生成、调整所属资源的访问控制策略,提高访问控制的灵活性与可扩展性.

(4) 区块链基于分布式共享总账技术能够有效保证策略制定来源可靠、存储可信,通过分布式节点的共识机制提高了访问控制系统的抗攻击能力,有效防止单点故障的发生,保证系统可用性.

(5) 区块链是一种只增不删的数据管理模式,事务数据永远存储在区块链,区块链上存储的数据无法被篡改,能够实现对共享和流通过程中大数据资源的全流程追踪管控,便于系统审计.

1 相关工作

当前针对大数据访问控制领域的研究还处于起步阶段,但在该方向国内外的研究已经取得了一定的研究进展.针对医疗大数据中安全管理员难以预测医生实际的数据访问需求,授权管理难度增加的问题,惠榛等[11]提出了面向医疗大数据的风险自适应的访问控制模型,该模型通过分析医生的访问历史,使用信息熵和 EM (Expectation Maximization)算法量化医生侵犯隐私的风险,以此来适应性地调整医生的访问能力,防止过度授权的发生.针对传统访问控制模型难以描述大数据所具有的时空属性、访问控制客体描述困难的问题,文献[12, 13]分别提出了基于位置感知的访问控制模型 LARB 与 GEO-RBAC,将用户空间位置信息引入基于角色的访问控制模型中,结合用户位置属性来授予用户相应访问控制权限.针对大数据环境存在海量角色难以进行权限管理的问题,文献[14-16]基于角色工程通过自上向下或自下而上的形式对用户角色进行挖掘,高效地为用户提供个性化服务,以此提高大数据环境下权限管理的效率.另外,为了实现对大数据中个人隐私数据的保护,以基于属性加密的访问控制为代表的密文访问控制技术,作为一种利用密文机制实现访问控制的方法,也得到了很多学者[17, 18]的广泛研究,但该技术还存在策略表达能力不足、安全性与效率性能不能兼顾等问题,在实际应用中还面临诸多挑战[19].

目前分布式访问控制的研究主要围绕分布式实体间跨域互操作的问题[20],一般采用角色映射或属性转换的方法实现域间权限转换.根据域间协作架构的不同,可分为联邦式架构和松耦合式架构[21, 22].联邦式架构是一种中心式架构,由单一授权中心与多个访问控制代理组成,通过授权中心设置面向分布式系统的全局角色来实现用户在不同域内权限的转换;松耦合式架构是一种多中心式架构,由多授权中心与多个访问控制代理组成,通过各域内的授权中心对域外用户权限进行转换.角色映射或属性转换机制需要提前进行协商,建立信任关系,仅适用于少量实体参与的有限分布式场景中,难以满足大数据环境高动态性、强灵活性的需求.

另外,当前基于区块链的访问控制研究渐渐兴起.如为了提高分布式访问控制策略管理的灵活性,Damiano

等[23]探索了使用基于区块链交易的形式来创建、管理、执行访问控制策略的可行性,并通过比特币平台进行了实现,但仅仅是将区块链作为策略管理的数据库,还需要第三方应用来提供集中式访问控制服务.Zyskind[24]基于区块链技术实现了移动应用程序的粗粒度权限管理,交易 T_{access} 用于管理策略,交易 T_{data} 用于存储和索引数据,区块链中每个用户和服务都对应一个公钥地址作为身份的凭证,可由用户公钥（资源拥有方）与服务公钥（资源请求方）共同以联合身份的形式对权限进行管理.针对物联网数据的访问控制问题,FairAccess[25-28]则将策略以(resource, requester)的形式存储在区块链交易中,引入比特币中 Wallet 概念,为不同的 IoT 设备安装自己的 Wallet,Wallet 起到访问控制代理的功能,通过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.针对医疗数据的访问控制问题,MedRec 框架[29,30]基于以太坊平台将智能合约与访问控制相结合进行自动化的权限管理,实现了对不同组织的分布式医疗数据的整合和权限管理,MedRec 框架包括三个层次的合约 Registrar Contract、Patient Provider Relationship、Summary Contract,但 MedRec 权限管理不够灵活且选择将策略直接存储在智能合约,随着策略规模的增大,智能合约的运行成本将变得十分高昂,不适用于大规模的动态授权应用场景.

综上分析可知,当前针对大数据访问控制的研究还处于起步阶段,缺少对访问控制分布式与动态性特点的考虑,而传统面向分布式系统访问控制的研究又难以适用于大数据环境.另外,目前基于区块链的访问控制技术还没有同时实现策略管理与访问控制决策的分布式.

2 预备知识

本节定义基于区块链的大数据访问控制机制使用到的预备知识,为后文阐述方便,主要对区块链技术及智能合约的基本概念进行简要介绍,并对基于属性的访问控制模型进行形式化定义.

2.1 区块链技术基本原理

区块链是一种在对等网络环境下,基于透明和可信共识规则并按照时间顺序将数据区块以链条的方式组合形成的特定数据结构,并以密码学方式保证其数据不可篡改、不可伪造、可追溯的去中心化、去信任的分布式共享总账系统.区块链的“分布式”不仅体现为数据备份存储的分布式,也体现在数据记录的分布式,即由所有节点共同参与数据维护,单一节点的数据被篡改或被破坏不会对区块链所存储的数据产生影响,能有效避免单点故障的发生.并且,由于大数据自身的特点,大数据的采集、存储、分析等同样是分布式的应用场景,同样需要进行信息及价值的交换.因此,区块链技术与大数据的访问控制需求具有高度的契合点.

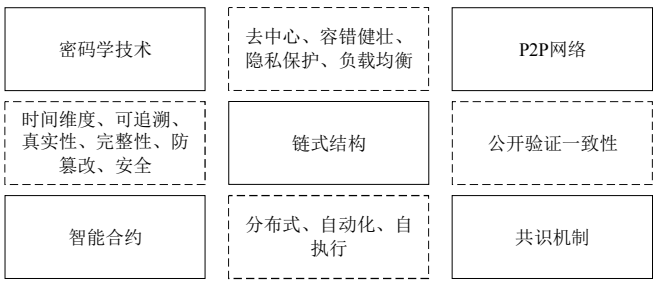

Fig.1 Blockchain technology features

图 1 区块链技术特点

区块链并不是单一的技术创新,而 P2P 网络技术、密码学技术、Merkle 树、共识机制、智能合约等多种技术深度整合的结果,能够通过透明和可信规则,实现事务的管理.如图 1 所示,加密的链式结构用来验证和存储数据,P2P 网络技术和共识机制用来实现分布式节点的验证和通信,智能合约能够实现复杂业务逻辑功能并对数据进行自动化操作.这些技术整合到一起,形成了一种新的数据记录、存储和表达的方法.

2.2 智能合约技术

智能合约[31, 32]是存储在区块链上能够在每个分布式网络节点上自动运行的脚本.1994年,由Nick Szabo首次提出智能合约的概念,定义其是一种“通过计算机执行合同条款的交易协议”,即通过代码程序来自动执行合同[33].只要满足合同条款,交易将无需第三方监督自动进行.虽然智能合约的概念很早就被提出,但由于缺乏支持可编程智能合约运行的信息平台及相关技术,智能合约一直停留在概念阶段,直到能够作为“信任机器”的区块链的出现才为智能合约技术的应用落地提供了平台支撑.由于区块链具有去中心化、安全、不可篡改、透明可追踪等优点,为智能合约提供了可信的执行环境.作为“区块链2.0”的核心特性,智能合约能够在去信任环境下,按顺序触发设定的合约内容并完成一系列安全的自动化操作.同时,区块链数据具有完备可追溯的属性,还可支持事后审计以追踪合约动态.第二大区块链平台以太坊[34]设计了一种基于“EVM虚拟机”的图灵完备脚本语言,极大的扩宽了区块链的应用领域.智能合约为区块链提供了应用层的扩展接口,任何开发人员都可基于底层区块链技术通过脚本实现其所要实现的工作,为区块链的应用落地奠定了基础.

2.3 基于属性的访问控制模型

针对传统访问控制模型难以解决的动态、细粒度访问控制问题,研究人员提出了基于属性的访问控制模型(Attribute-Based Access Control, ABAC).ABAC模型基于实体属性而不是用户身份来判决允许或拒绝用户对资源的访问控制请求.ABAC 模型的核心要素包括主体、资源、操作以及环境约束,这些要素统一使用属性和属性值来进行表示,属性间的关系可以根据访问控制需求进行灵活的设置,提高了访问控制策略语义的表达能力和模型的灵活性,并且能够将其它访问控制模型中权限、安全标签、角色等概念用属性来进行统一描述,适用于解决分布式环境下动态大数据的访问控制问题.基于如下原因,我们认为 ABAC 模型相比其它模型能更好的适用于大数据访问控制场景.

(1)细粒度访问控制:ABAC模型通过属性来对实体及约束进行描述,能够严格控制访问者取得权限的各种条件,精确设定属性-权限关系,实现最小权限原则.

(2)自主授权:ABAC 模型可为资源拥有者提供策略管理接口,策略无需由管理员统一设定,资源拥有者可以根据自身实际资源保护需求发布、更新、撤销策略,保证资源能够按照资源拥有者的意愿被访问.

(3)动态访问控制:ABAC模型依据请求者所具有的属性集合决定是否赋予其访问权限,实现了策略管理和权限判定的分离,且属性的设置与更新具有极大的灵活性和扩展性,可满足不同应用场景需求.

(4)较小的系统开销:在用户和资源数量大幅度增加的情形下,传统 DAC、RBAC 等访问控制模型策略数目将呈指数级增长,系统维护难度及开销将大大增加.而 ABAC 模型中,策略随用户和资源的增长呈线性增加,当达到一定规模后,系统开销趋于平稳[35].

为了便于本文的叙述,给出如下定义.

定义 1 属性项(Attribute Item):是表示属性的基本单元,用{$xAttrName=attrValue$}, ($xAttrName \in attrSet$, $attrValue \in$ Range($xAttrName$), $x \in \{s, r, a, e\}$)表示,$xAttrName$ 表示属性名,$attrValue$ 表示属性值,为了对不同类型的属性表示方便,用 x 表示属性类型,s、r、a、e 分别代表主体属性、资源属性、动作属性和环境属性.

定义 2 属性元组(Attribute Tuple):是同类型属性项的集合,用 $xAttrTuple$, $x \in \{s, r, a, e\}$表示,即 $xAttrTuple$: $\{(xAttrName_1=attrValue_1) \wedge (xAttrName_2=attrValue_2) \wedge \ldots \wedge (xAttrName_i=attrValue_i)\}$.

定义 3 属性访问请求(Attribute Access Request, AAR):由一组主体属性、资源属性、动作属性和环境属性组成,用 AAR: {$sAttrTuple \wedge rAttrTuple \wedge aAttrTuple \wedge eAttrTuple$}来表示,AAR的含义是属性为 $sAttrTuple$ 的请求者在环境属性 $eAttrTuple$ 下对资源 $rAttrTuple$ 请求进行操作 $aAttrTuple$.

定义 4 访问控制策略:针对资源的访问控制规则,体现了资源拥有者的授权行为,规定了访问受保护资源所需要具有的属性集合,记为 Policy: result(R, $action$, pid)$\leftarrow \Theta\{xAttrTupleSet\}_{signature\_owner}$, $x \in \{s, r, a, e\}$.其中,$\Theta\{xAttrTupleSet\}$表示由属性项集合 $xAttrTupleSet$ 中的属性通过合取、析取等逻辑关系构成的逻辑表达式,pid 表示策略ID.当请求方所拥有的属性使 $\Theta\{xAttrTupleSet\}$为真时,请求方能够被允许或拒绝对资源 R 进行 $action$

操作,*result*∈{Permit, Deny}.另外,策略需要被资源拥有者或策略发行方签名后在区块链中保存,从而保证发布策略的真实性.

3 BBAC-BD 机制框架与工作流程

3.1 大数据访问控制架构

大数据访问控制涉及大数据资源的采集、汇聚、管理、控制等.大数据访问控制架构主要由数据层、资源汇聚层、基础设施层、事务层、共识层、访问控制合约层 6 部分（如图 2 示）组成.各层结构相互协同又各司其职,共同构成一个完整的大数据访问控制架构.

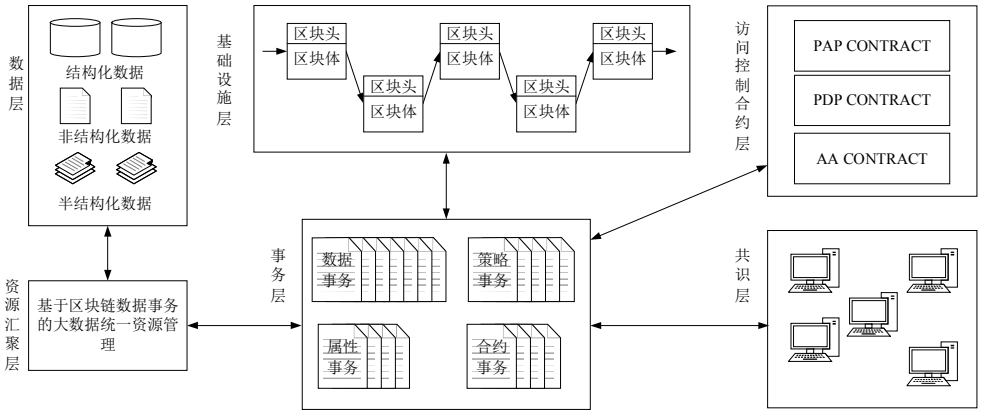

Fig.2 Big data access control technology architecture

图 2 大数据访问控制技术架构

（1）数据层:真实的大数据资源,包括结构化数据、非结构化数据和半结构化数据,分布式的存储于不同的位置,逻辑上受资源汇聚层的统一管理.

（2）资源汇聚层:基于区块链技术来对大数据资源进行资源管理,实现不同来源大数据资源的汇聚.虽然,真实的大数据资源在实际上是由不同的数据拥有方分布式的存储,但是通过区块链技术在逻辑上形成对大数据资源的统一管理.本文的研究重点是访问控制机制,资源汇聚层不做详细的阐述.

（3）基础设施层:由区块链平台为大数据访问控制提供基础设施,是整个架构的基础,需要全网节点、矿工来维持系统的正常运行,是大数据访问控制平台事务和智能合约的载体.事务层、合约层、资源汇聚层都是以区块链为基础的上层应用.

（4）事务层:包括数据事务、策略事务、属性事务、合约事务四种类型的访问控制类事务.数据事务用于对大数据资源进行管理,服务于资源汇聚层;策略事务用于对访问控制策略的管理,包括策略的发布、更新和撤销,为合约层的 PAP CONTRACT 提供数据支撑;属性事务用于对实体属性的管理,包括属性的发布、更新和撤销,为合约层 AA CONTRACT 提供数据支撑;合约事务用于为智能合约提供运行环境,服务于合约层.

（5）共识层:主要包括共识机制,通过各类共识算法来保证分布式节点间访问控制数据的一致性和真实性,从而在节点间达成稳定的共识.

（6）访问控制合约层:包括 PAP CONTRACT、PDP CONTRACT、AA CONTRACT 三种合约.PAP CONTRACT 用于访问控制策略管理,PDP CONTRACT 用于访问控制请求判决,AA CONTRACT 用于实体属性管理.

3.2 BBAC-BD框架及工作流程

本节对基于区块链的大数据访问控制机制 BBAC-BD 的基本框架进行详细的说明,介绍该机制的工作流程.本文提出的基于区块链的大数据访问控制框架如图 3 所示,框架基于 ABAC 模型进行了改进,将区块链技术

与访问控制技术相结合.框架包括策略执行点(Policy Enforcement Point, PEP)、属性权威(Attribute Authority, AA)、策略管理点(Point Administration Point, PAP)、策略决策点(Policy Decision Point, PDP)四个核心部分.其中,AA、PAP、PDP 用智能合约的方式来实现.为了能够确保区块链中访问控制策略的正确执行,用户需要使用 PEP 作为访问控制客户端来与区块链进行访问控制的交互.

BBAC-BD 框架中访问控制工作流是对标准 ABAC 模型工作流的扩展.访问控制工作流程可分为准备阶段、执行阶段两个阶段（如图 3 示）.准备阶段主要进行访问控制策略及属性的管理,包括策略及属性的发布、更新、撤销以及对策略与属性查询结果的响应.而执行阶段主要进行访问请求的判决、响应与执行.

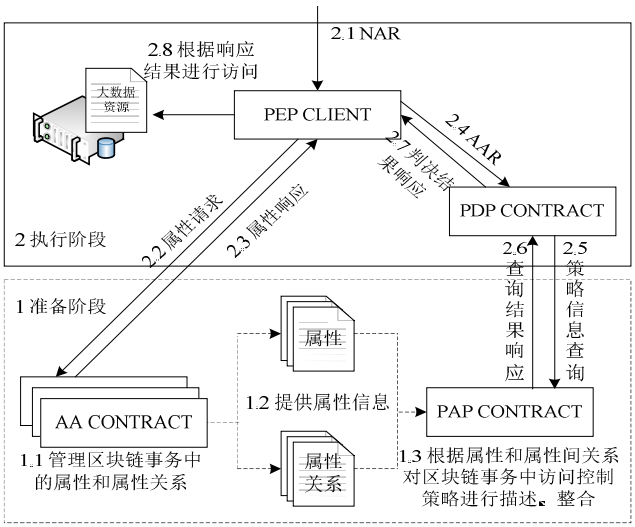

Fig.3 BBAC-BD framework

图 3 BBAC-BD 框架

准备阶段:1）由属性发布方向区块链中发布属性及属性关系信息,由 AA CONTRACT 预先收集、整合区块链事务中属性信息,以供 PEP CLIENT 和 PAP CONTRACT 使用;2）由策略发布方向区块链中发布访问控制策略,由 PAP CONTRACT 结合属性信息描述、收集、整合区块链事务中访问控制策略,以供 PDP CONTRACT 进行访问请求的判决.

执行阶段:1）当 PEP CLIENT 收到用户向其发送对某一资源执行某项操作的请求时,PEP CLIENT 分析得到原始访问请求中主体、客体和操作语义,根据从 AA CONTRACT 得到的属性信息生成基于属性的访问请求 AAR,将 AAR 发往 PDP CONTRACT.2）PDP CONTRACT 向 PAP CONTRACT 查询与被请求大数据资源相关的访问控制策略集,进行访问控制判决,将判决结果响应发送回 PEP CLIENT.3）由 PEP CLIENT 根据响应结果对大数据资源进行授权的访问操作.

由于访问控制策略是存储在区块链中,策略信息对任何人都是可验证、可追溯且不可篡改的,大数据资源的访问控制摆脱了传统集中式访问控制管理可能存在的单点故障和访问控制判决透明度的问题,实现了访问控制策略的分布式管理,有效地提高了系统的鲁棒性和可信性.另外,通过智能合约的形式来实现访问控制策略的判决过程,无需第三方中心机构参与,避免了第三方中心可能存在的越权行为,基于区块链系统实现共识下的访问控制自动判决,是一种真正实现了去中心化的访问控制机制,符合大数据资源的访问控制管理需求.

4 面向 BBAC-BD 策略管理的事务结构

4.1 区块链中事务存储结构

在区块链中,数据以事务的形式存储在区块链中,我们使用区块链中事务的形式来对访问控制策略进行管理.如图 4 所示,区块中事务数据是以基于哈希算法的 Merkle 树这种数据结构进行存储,通过哈希算法将大小不

一致的事务数据映射成固定大小的字符串,存储在 Merkle 树的叶子节点上,Merkle 树的非叶子节点存储的都是其子节点的哈希值.在 P2P 网络中,使用 Merkle 树能够快速的验证数据是否被篡改或接收到的数据是否损坏,区块链中所有的事务数据通过 Merkle 树生成唯一的 Merkle 根存储在区块头中.

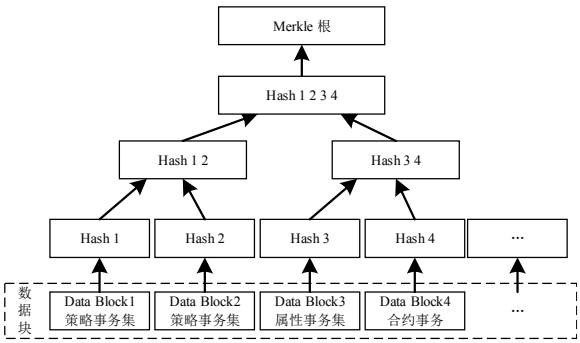

Fig.4 Merkle tree storing transaction information

图 4 存储事务信息的 Merkle 树

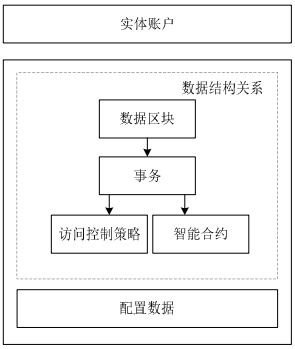

Fig.5 Entity relationship in the block

图 5 区块中数据视图实体间关系

区块链包括实体账户、数据区块、事务集数据、配置数据等,它们间数据视图关系如图 5 所示.

1)实体账户:是区块链中各类事务请求的发起者及相关数据的拥有者,是访问控制机制中数据资源的实际拥有者,拥有一对由 PKI 体系产生的公钥与私钥,系统中账户由公钥唯一进行标识,由资源拥有者发布的事务数据都需由实体账户用私钥对其进行签名,以供其它用户验证区块中事务的真实性.

2)数据区块:是区块链网络中底层的数据,多个区块共同形成链式结构以不可篡改的形式将一定时期内的事务处理结果持久化.

3)事务集数据:是基于区块链的访问控制机制中存储执行实际业务活动的数据,包括访问控制类事务和智能合约类事务.访问控制类事务用于访问控制策略的管理,主要涵盖策略管理中策略信息的发布、更新与撤销操作.智能合约类事务用于访问控制的判决过程,响应访问请求,生成访问控制响应.在区块链上发布事务需要消耗一定数据的代币,作为用户使用区块链服务的开支.

4)配置数据:是区块链系统正常运行所需的配置信息,包括协议版本号、通信节点信息等配置信息.

4.2 基于事务的访问控制策略管理

我们以区块链中事务为载体来对访问控制策略进行管理.访问控制策略管理包括策略信息管理和属性信息管理,分别以策略类事务和属性类事务的形式发布到区块链,包括信息的发布、更新与撤销操作.策略管理信息由大数据资源的拥有者制定,对大数据资源的访问控制策略权限信息和属性及属性间关系进行管理.数据被发布到区块链后,可以被任意次更新,直到数据被撤销,才结束策略或属性信息的完整生命周期.无论数据当前处于何种状态(发布、更新、撤销),其历史操作信息都将被永久记录在区块链中,便于审计和掌握访问控制动态.区块链非合约事务的消息格式如图 6 示.

ID	PK	transactionType	action	transaction _ data	timestamp	signature_message

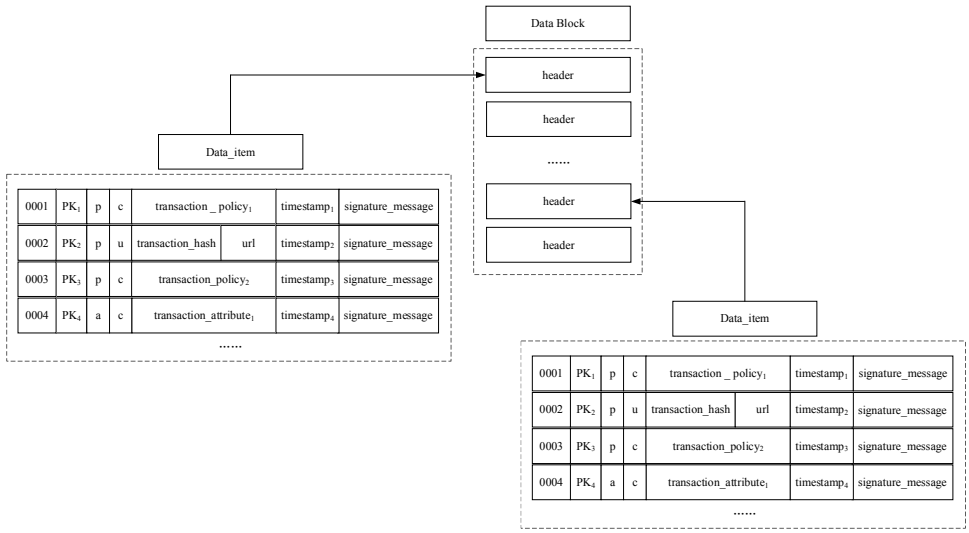

Fig.6 Transaction management level relationship

图 6 事务管理层级关系

其中,ID 表示事务消息标识号;PK 表示事务发布者公钥;transactionType 代表事务消息类型,用 p 表示策略类事务消息,a 表示属性类事务消息;action 代表操作类型,用 c 表示发布操作,u 表示更新操作,r 表示撤销操作;transaction_data 代表具体的事务消息数据,与事务消息类型相对应,分别访问控制策略信息和属性及属性关系信息,根据事务消息数据规模的不同,存在链上与链上链下相结合两种存储方法,针对小规模策略数据,直接在链上存储策略信息,针对较大规模策略数据,在链上存储数据摘要 Hash 和链下数据链接 Url;timestamp 表示消息发布的时间戳;signature_message 表示对前 4 项事务数据的消息签名.对于发布操作,需要消耗一定数量的数字代币,更新与撤销操作无需消耗数字代币.

下面是矿工节点接收事务数据生成新区块的具体工作过程:

步骤 1:策略信息发布者向区块链提交策略事务请求,并且使用发布者公钥作为事务标识.

步骤 2:由代表矿工节点接受策略信息请求,并向节点网络广播所接受到的信息请求.

步骤 3:由当值的代表矿工节点根据用户的公钥将事务记录到 child_block.

步骤 4:由当值的代表矿工节点将 child_block 的确认信息向节点网络进行广播.

步骤 5:校验代表矿工节点对事务数据进行校验,其它矿工节点同步更新事务数据.

步骤 6:每隔一个时间时隔对 child_block 中的策略事务数量进行检查,当数据达到 5 个时,将所有等更新事务数据打包封装成一个数据区块并计算该区块的根 Merkle 值,并将该区块数据存入本节点数据库.

步骤 7:返回步骤 1.

5 面向 BBAC-BD 的智能合约

5.1 策略管理合约PAP CONTRACT

访问控制策略以事务的形式存放在区块链中,并且每个事务中可能存在一条或多条访问控制策略.PAP 需要对区块链中存储的策略事务进行整合,将与访问请求相关的策略发送给 PDP,以供 PDP 进行策略判决.如图 7 所示,为访问控制策略的融合流程,图 7 中在区块链存储的每条策略信息的第 1 个操作类型字段表示了该策略信息被发布的意图,c 表示发布操作,u 表示更新操作,r 表示撤销操作.并且每条策略信息都有对应的时间戳为该信息增加了时间维度,根据时间维度,当操作类型字段为 c 时,创建相应 ID 的新策略信息,当操作类型字段为 u 时,更新相应 ID 的策略信息,当操作类型字段为 r 时,撤销相应 ID 的策略信息.访问控制策略的融合流程即是沿着相应 ID 策略的时间维度管理整合策略的过程.策略管理合约 PAP CONTRACT 将 AAR 所请求资源相关策略

的进行整合,为 PDP CONTRACT 进行访问控制判决提供策略支撑.

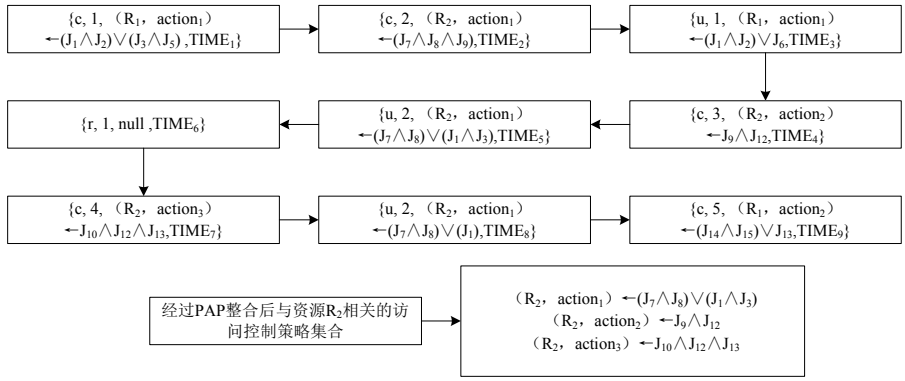

Fig.7 Access control policy fusion process

图 7 访问控制策略融合流程

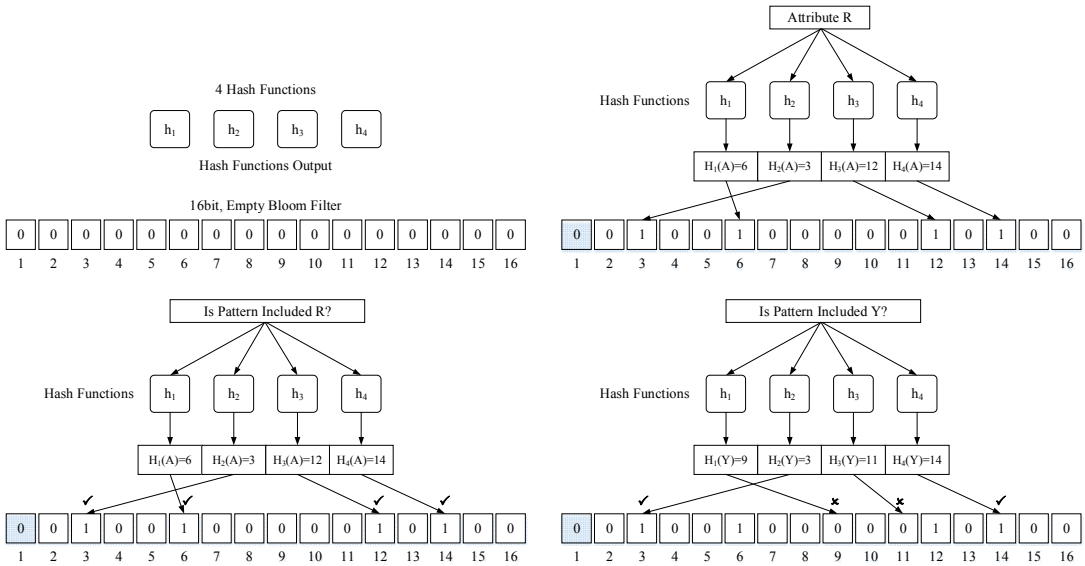

Fig.8 Access control policy management based on Bloom Filter

图 8 基于 Bloom Filter 访问控制策略管理

访问控制策略分布式存储后面临的关键问题是策略查询效率的问题.为了提高策略管理过程中策略检索与查询的效率,本文设计了基于 Bloom Filter 的策略管理合约.作为一种具有极高空间利用效率的概率性数据结构,Bloom Filter 通过二进制向量来描述数据集合,能够快速判断该集合中是否包含某一特定元素.Bloom Filter 的缺点是存在一定的错误查询概率,Bloom Filter 不会把集合中存在的元素判断为不存在,但存在把集合中不存在的元素判断为存在集合中的可能.Bloom Filter 正是通过允许存在少量的错误,以此来减少存储空间,提高查询效率.在策略管理合约中,允许非相关策略被判定为相关策略,这里该非相关策略将成为冗余策略,而不允许相关策略被判定为非相关策略,这与 Bloom Filter 的特点正好吻合.

Bloom Filter 中包括长度为 n 比特的二进制向量 BF={b_0, b_1, … , b_{n-1}}与 m 个独立的哈希函数 H(x)={ $h_0(x)$, $h_1(x)$, … , $h_{m-1}(x)$},初始化阶段将 BF 的所有比特位都设定成 0,通过 H(x)的计算可以得到于 0~n-1 范围内分布均匀的哈希值.对于原始策略事务数据集合中的属性关键字集合 DB_SET={DB_0, DB_1, … , DB_{k-1}},将属性关键

字 DB_i ($0\le i\le k$)插入到 BF 中时,计算与 m 个哈希函数 H(x)对应的 m 个哈希值 $h_0(DB_i)$, $h_1(DB_i)$, … , $h_{m-1}(DB_i)$,并将在 BF 中 m 个哈希值所相对应位置的值设定成 1.当需要验证某一资源的相关策略属性关键字 R 是否存在于该 BF 中时只需要计算 $h_0(R)$, $h_1(R)$, … , $h_{m-1}(R)$,并在查看 BF 中 $h_0(R)$, $h_1(R)$, … , $h_{m-1}(R)$对应位置的值是否全部为 1,若不全部为 1,则证明该 BF 中涉及的资源数据集合不包含资源 R 的相关策略.若对应位置的值全部为 1,则认为资源 R 相关策略以 $(1-P_R)$的概率存在于该事务数据集合,m_{op} 为最优哈希个数.

$$P_R=(1-(1-\frac{1}{n})^{m\cdot k})^m\approx(1-e^{\frac{-m\cdot k}{n}})^m$$

$$m_{op}=\ln 2(\frac{n}{k})$$

PAP CONTRACT 合约的伪代码如下所示.

算法 1. 策略管理合约 PAP CONTRACT

输入:属性访问请求 AAR,区块链 blocks

输出:资源相关策略集 RELEVANT_POLICY_SET

```
1.   rAttrTuple = attributeParser(AAR); this = currentBlock; mark = null;
2.   for i = 1 to blocks.length do
3.        {for j = 1 to this. policy_datablock.length do
4.             {BF = getBloomFilter(this.policy_datablock[i]); mark =1;
5.              for k = 1 to this.datablock.hashfunction.length do
6.              {key[k] = hash (k, rAttrTuple);
7.              if (BF.NotContain(key[k])) mark = 0;}
8.              if (mark = 1)   then
9.              {RELEVANT_POLICY_SET. increase(this.datablock.transaction.policy); continue;}
10.             else continue;
11.            }   //end for
12.       }   //end for
13.  return RELEVANT_POLICY_SET;
```

PAP CONTRACT 算法流程描述:

1）解析 AAR 得到所请求的资源属性信息 rAttrTuple;

2）遍历区块链各区块内策略类事务数据块;

3）获取策略类事务数据块所对应的属性布隆过滤器 BF;

4）根据该 BF 所对应的哈希函数,计算 rAttrTuple 所对应的哈希值;

5）若计算 rAttrTuple 得到的哈希值与 BF 中对应位全部为 1,则在 RELEVANT_POLICY_SET 中添加该数据块内策略,否则,该数据块中无 rAttrTuple 的相关策略;

6）当对所有区块内策略类事务遍历完成后,向 PDP 返回资源相关策略集 RELEVANT_POLICY_SET 用于权限判决.

5.2 策略判决合约PDP CONTRACT

访问控制的判决结果分为两种类型,分别是允许访问(PERMIT)和拒绝访问(DENY).针对 PDP 收到的 AAR,若该 AAR 满足某一访问控制策略中的约束和谓词,则此判决请求为满足策略,根据策略描述来 PERMIT(PID)或 DENY(PID)该 AAR 请求.如果判决请求不满足策略,则包括两种情况,一是 AAR 中所提供的请求属性信息不足,从而无法做出判决,使用表示 UNKNOWN,另一类是策略集没有任何一条策略能够与 AAR 进行匹配,从而无法做出判决 UNSATISFY.因此,访问控制策略判决阶段包括四类判决结果 PERMIT、DENY、UNKNOWN 和 UNSATISFY.

AAR所包括的全部属性构成了访问请求的关联属性集,用$ATTR_SET_{AAR}$来表示.访问控制策略policy中所包括的全部属性构成了访问控制策略的关联属性集,用AAT_SET_{policy}来表示.针对特定的访问请求,访问控制的判决PolicyDecide可以表示成如下形式:

{$ATTR\_SET_{AAR}$, AAT_SET_{policy}}->{ PERMIT、DENY、UNKNOWN、UNSATISFY}

PolicyDecide流程如下:

1）若$AAT_SET_{policy} \not\subset ATTR_SET_{AAR}$,PolicyDecide(policy)=UNKNOWN;

2）若$AAT_SET_{policy} \subseteq ATTR_SET_{AAR}$,且$ATTR_SET_{AAR}$中属性值范围全部符合$AAT_SET_{policy}$中策略属性约束要求,则 *PolicyDecide*(policy)=SATISFY,SATISFY∈{ PERMIT, DENY}.

3）若$AAT_SET_{policy} \subseteq ATTR_SET_{AAR}$,且$ATTR_SET_{AAR}$中存在至少一个属性值不符合$AAT_SET_{policy}$中策略属性约束要求,则PolicyDecide(policy)=UNSATISFY.

策略判决过程由策略判决合约PDP CONTRACT来进行,合约的伪代码如下所示.

算法2. 策略判决合约PDP CONTRACT

输入:属性访问请求AAR,访问控制策略集POLICY_SET

输出:策略判决结果PERMIT、DENY、UNKNOWN、UNSATISFY

```
1.  UNKNOWN_SET = POLICY_SET; PERMINT_RESULT_SET = null;
2.  DENY_RESULT_SET = null; UNSATISFY _RESULT_SET = null;
3.  for i = 1 to UNKNOWN_SET.length do
4.      {result = PolicyDecide(UNKNOWN_SET[i]);
5.      if (result = permit) then
6.          { UNKNOWN_SET.delete(UNKNOWN_SET[i]);
7.          PERMINT_RESULT_SET.add(UNKNOWN_SET[i].PID);}
8.      elseif (result = deny) then
9.          { UNKNOWN_SET.delete(UNKNOWN_SET[i]);
10.         DENY_RESULT_SET.add(UNKNOWN_SET[i].PID);}
11.     elseif (result = unsatisfy) then
12.         { UNKNOWN_SET.delete(UNKNOWN_SET[i]);
13.         UNSATISFY _RESULT_SET.add(UNKNOWN_SET[i].PID); }
14.     }   //end for
15. if (PERMINT_RESULT_SET ≠ null && DENY_RESULT_SET == null) then
16.     return PERMIT;
17. elseif (DENY_RESULT_SET ≠ null && PERMINT_RESULT_SET == null) then
18.     return DENY;
19. elseif (PERMINT_RESULT_SET ≠ null && DENY_RESULT_SET ≠ null) then
20.     return conflict_handle();
21. else
22.     return UNKNOWN;
```

PDP CONTRACT算法流程描述:

1）将所有待判决策略放入UNKNOWN_SET,作为预判决策略集;

2）遍历预判决策略,分别得到四个策略判决结果集PERMINT_RESULT_SET、DENY_RESULT_SET、UNSATISFY_RESULT_SET、UNKNOWN_RESULT_SET.

3）根据判决结果集,得出AAR请求的最终判决结果,若存在冲突的判决结果,进行冲突处理后,得到最后判决结

果.冲突处理可依据肯定优先或否定优先等处理原则进行冲突消解.

5.3 属性权威合约AA CONTRACT

PEP、AAR、PDP 中属性语义与属性值赋值都来源于属性权威 AA.系统中可能存在多个 AA,AA 中存储主体属性、资源属性、动作属性和环境属性的属性值和属性间关系的列表.如表 1 所示,左侧数据表示相关属性的属性值,右侧表示相关属性间的属性关系.

Table1 Attribute and attribute relationship

表 1 属性及属性关系

sID(0001) sID(0002) sID(0003) sID(0004) sName(John) sName(Robert) sName(Alice) sName(Bob) sRole(full professor) sRole(associate professor) sRole(instructor) sRole(others) rID(00001) rID(00002) rID(00003) rName(transcript data) rName(paper data) rName(project data)	sIdNameRelavant(0001,John) sIdNameRelavant(0002,Robert) sIdNameRelavant(0003,Alice) sIdNameRelavant(0004,Bob) rIdNameRelavant(00001,transcript data) rIdNameRelavant(00002,paper data) rIdNameRelavant(00003,project data) roleAssignment(0001,full professor) roleAssignment(0002,associate professor) roleAssignment(0003,instructor) roleAssignment(0004,others) allowAction(transcript data, create) allowAction(transcript data, read) allowAction(transcript data, update) allowAction(transcript data, delete) allowAction(paper data, create) allowAction(paper data, read) allowAction(project data, create) allowAction(project data, read) allowAction(project data, delete)

属性信息同样是以事务的形式存储在区块链中,基于 Bloom Filter 进行属性管理,AA CONTRACT 相当于提供 AA 查询服务的代理,合约的伪代码如下所示.

算法 3. 属性权威合约 AA CONTRACT

输入:属性类事务 atribute_transaction,属性请求 attributeRequest

输出:相关属性集 RELEVANT_ATTRIBUTE_SET

```
1.  AttrTuple = attrbuteRequestParser(attributeRequest); this = currentBlock; mark = null;
2.  for i = 1 to blocks.length do
3.      {for j = 1 to this. attribute_datablock.length do
4.          {BF = getBloomFilter(this.attribute_datablock[i]); mark =1;
5.           for k = 1 to this.datablock.hashfunction.length do
6.           { key[k] = hash (k, rAttrTuple);
7.           if (BF.NotContain(key[k])) mark = 0;
8.           }   //end for
9.           if (mark = 1) then
10.          {RELEVANT_ATTRIBUTE_SET.increase(this.datablock.transaction.attribute); continue;}
11.          else continue;
12.         }   //end for
13.     }   //end for
14. return RELEVANT_ ATTRIBUTE _SET
```

AA CONTRACT 算法流程描述:

1）接收属性查询请求 attributeRequest;

2）遍历区块链各区块内属性类事务数据块;

3）获取属性类事务数据块所对应的属性布隆过滤器 BF;

4）根据该 BF 所对应的哈希函数,计算对应属性的哈希值;

5）若计算得到的哈希值与 BF 中对应位全部为 1,则在 RELEVANT_ATTRIBUTE_SET 中添加该数据块内属性信息,否则,该数据块中无相关属性信息;

6）当对所有区块内属性类事务遍历完成后,将 RELEVANT_POLICY_SET 作为响应进行返回.

6 实验仿真与分析

通过仿真实验对本文所提出的 BBAC-BD 机制有效性进行测试,以验证 BBAC-BD 机制是否能够实现访问控制策略管理与访问控制策略判决功能.本文基于 XACML 提供的标准策略一致性测试包中属性集和策略集进行测试,1-6 组的 Policy Sample 分别与 1000、2000、3000、4000、5000、8000 条单一策略测试集样本对应,对开源区块链平台 EbCoin 进行了扩展与改进,改用 PoW 机制,将 PAP CONTRACT、PDP CONTRACT、AA CONTRACT 三部分合约代码与 EbCoin 进行整合在 PC 机上构建三节点仿真实验环境.实验环境如下:操作系统为 Windows 10 家庭中文版 64 位,CPU 为 Intel(R) Core(TM) i7-4710MQ @ 2.50GHz,内存大小为 16.00GB,nodejs 版本为 v8.11.1,npm 版本为 5.6.0.

6.1 策略检索效率测试

本实验针对基于 Bloom Filter 策略管理方法对策略检索效率的优化效果进行性能测试,实验分别针对不同查询规模的测试集进行匹配查询测试,为 Bloom Filter 设置不同参数进行误差率与检索时延的测试,用于评估 Bloom Filter 不同参数对检索性能的影响.单次时延的计算方法是总时延/总的匹配次数.优化效果主要通过策略管理合约中策略的检索时延进行衡量,时延越小,其执行效率越高,检索优化的效果越好.

Table2 Policy retrieval test results

表 2 策略检索测试结果

序号	n/k	m 测试值	m 最优值	查询规模	误判率（真实）	误判率（理论）	总时延（毫秒）	单次时延（毫秒）
1	20	3	14	4000	0.1957%	0.2702%	433	0.108
2	20	3	14	8000	0.1896%	0.2702%	797	0.099
3	20	6	14	4000	0.0312%	0.0303%	761	0.190
4	20	6	14	8000	0.0289%	0.0303%	1485	0.186
5	20	14	14	4000	0.0047%	0.0067%	986	0.247
6	20	14	14	8000	0.0059%	0.0067%	2061	0.258
7	20	20	14	4000	0.0092%	0.0104%	1248	0.312
8	20	20	14	8000	0.0089%	0.0104%	2461	0.308
9	10	3	7	4000	0.9365%	1.7405%	483	0.121
10	10	7	7	4000	0.729%	0.819%	1132	0.283
11	5	3	3	4000	9.11%	9.2%	838	0.209
12	2	1	1	4000	39.15%	39.3%	761	0.190
13	2	2	1	4000	39.88%	40%	810	0.202

由表 2 测试结果可知,n/k 的比值越大优化效果越好.比值越大,所对应的误判率会越低,但同时也会占用更多的空间成本.同时,误判率的真实值一般小于误判率的理论值.图 9 与图 10 结果表明,当 Hash 的性能较好,即 Hash 分布结果较均匀时,在 m=3 的条件下,就能够达到预期使用所能接受的误判率,Hash 次数的增加并不会带来明显的收益增加.因此,在条件允许的情况下,尽量扩大 n/k 的值,能够有效提高查询性能.这主要是由于 Bloom Filter 在策略检测的属性关键字过滤过程中,其检索时间受策略集合规模影响较小,故相比于遍历检索过程,基于 Bloom Filter 的策略检索与匹配能够达到较高的性能,基于 Bloom Filter 的策略检索能够有效节约缓存空间,减少对缓存的请求次数,提升策略查询效率以及策略管理业务隔离性.

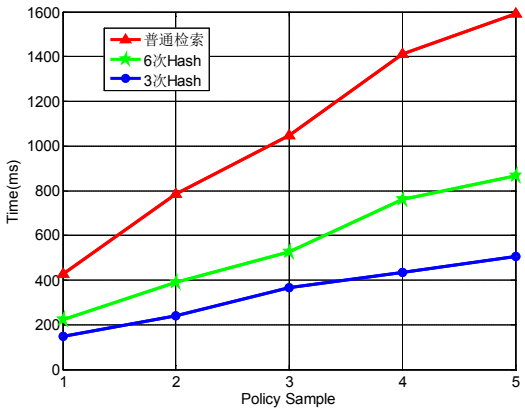

Fig.9 Policy retrieval performance

图 9 策略检索性能比较

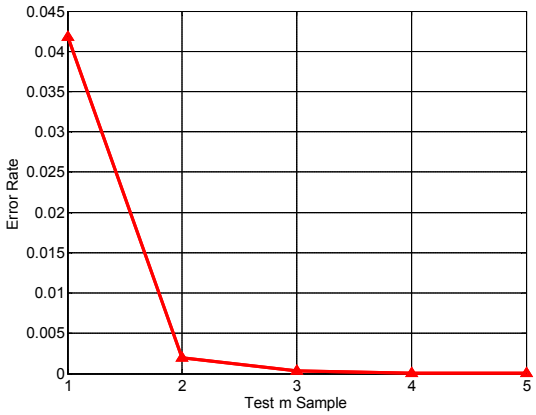

Fig.10 Effect of m value on the accuracy

图 10 m 值对检索结果准确性的影响

6.2 策略判决功能测试

为了验证 BBAC-BD 访问控制机制的有效性,本文在不同策略规模下对基于智能合约的访问控制策略判决功能进行了功能性测试,测试内容包括策略判决的效率与判决结果的成功率.1-5 组的 Policy Sample 分别与 1000、2000、3000、4000、5000 条单一策略测试集样本对应.测试集样本面向 80 个用户标识构建 800 次不同的访问请求,每个标识平均拥有 5 个属性值,每个请求共随机发送 5 次,策略判决时延通过计算所有请求的平均响应时延得到.

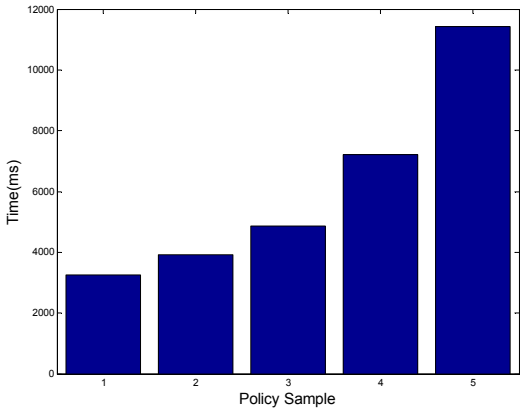

Fig.11 Policy decision performance

图 11 策略判决性能比较

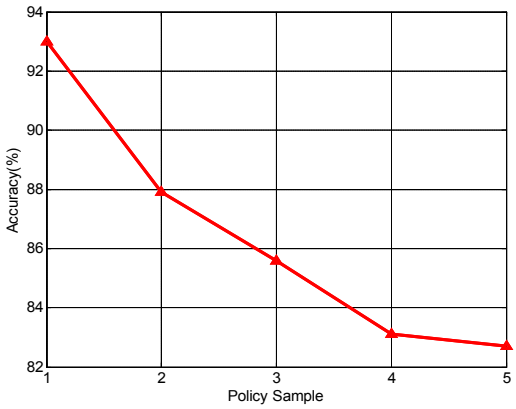

Fig.12 Policy decision success rate

图 12 策略判决结果成功率比较

由图 11 可知,策略判决时延与策略规模直接相关,随着策略规则的增加,访问控制判决时延增长较为明显.同时,由图 12 可知,随着策略规则的增加,策略判决成功率有所下降,这是由于策略集中存在部分冲突策略,针对冲突策略,策略判决合约无法得到一致性的判决结果,需要说明的是,本文还未将策略冲突处理部分引入策略判决合约,冲突策略的消解还未有效解决,这部分内容将在后续研究工作中继续完善.

6.3 区块链安全性分析

由于区块链所面临的主要安全风险来源于攻击者对共识机制的攻击,以此来达到修改区块数据的攻击目标.为了对区块链自身抗攻击安全性进行分析,以采用普遍的 PoW 共识机制为例,采取文献[36, 37]提出的攻击模型来分析区块链所面临的潜在安全风险.诚实节点可信链与恶意节点攻击链间的竞争关系可以用 Binomial Random Walk 过程来进行描述,攻击者伪造的攻击链长度成功超过可信链长度,从而弥补 z 个区块差距可能性

的问题可被近似的当成 Gambler's Ruin Problem.所以,攻击者成功弥补 z 个区块差距,成功完成对区块链数据篡改攻击的概率计算方法如下:

$$q_z=\begin{cases}1, p\le q\\ (\frac{q}{p})^z, p>q\end{cases}$$

式中 p 表示诚实节点获得下一区块记账权的概率,q 表示攻击者获得下一区块记账权的概率,且 p+q=1,q_z 表示攻击者最终成功弥补 z 个区块差距的概率.我们假设诚实节点以平均预期时间生成一个新区块,攻击者潜在区块链延伸长度符合泊松分布,其期望值如下:

$$\lambda=z\cdot\frac{q}{p}$$

为了计算攻击者所生成的区块链长度追赶上诚实节点所生成区块链长度的概率,将攻击者所生成区块长度的泊松分布概率密度与该时刻攻击者能够成功追赶诚实节点可信链的概率相乘,即为攻击者成功篡改区块数据的概率 p_a 为:

$$p_a=\sum_{k=0}^{\infty}\frac{\lambda^k\cdot e^{-\lambda}}{k!}\cdot\begin{cases}(\frac{q}{p})^{(z-k)}, k\le z\\ 1, k>z\end{cases}=1-\sum_{k=0}^{z}\frac{\lambda^k\cdot e^{-\lambda}}{k!}\cdot(1-(\frac{q}{p})^{(z-k)})$$

通过 Matlab 进行仿真,分析攻击者成功篡改区块数据的概率 p_a 与区块差距 z 及攻击者获得下一区块记账权的概率 p 之间的关系如图 13 所示,由仿真结果可知攻击者成功篡改区块数据概率随着区块链距离的增加呈现指数下降趋势,并且当区块距离相同时,成功篡改区块概率随着攻击者攻击能力的提升显著增加,当攻击者获取区块链网络内 50%以上的夺取记账权能力时,才能够控制整个区块链全部数据.当攻击者所掌握的夺取记账权能力较低时,通过采取适当的区块距离,区块链能够达到较好的抗攻击效果.

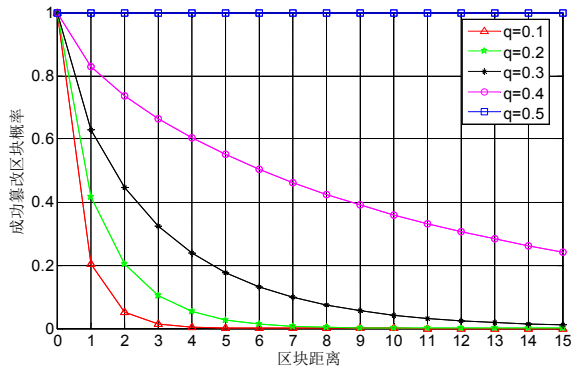

Fig.13 Attacker success probability

图 13 攻击者成功概率

7 结束语

本文提出了一种面向大数据资源的访问控制机制 BBAC-BD,该机制将区块链技术与 ABAC 模型相结合,借助区块链所具有的可追溯、不可篡改等特点,通过区块链事务管理访问控制策略及属性,实现了策略发布、更新以及撤销全流程的策略管理与追踪,策略以公开、透明的形式存放在区块链中,任何用户都可以对其进行查询,从传统基于第三方提供访问控制服务的模式中解脱出来,解决了权限判决透明度的问题.同时,通过智能合约基于资源拥有者发布到区块链上的策略实现对大数据资源自动化的访问控制,判决过程更加灵活、判决结果更加可信.BBAC-BD 机制实现了安全、可靠、透明的新型访问控制架构,能够有效促进大数据的安全流通与共享.

References:

[1] Dimitrov D V. Medical Internet of Things and Big Data in Healthcare. Healthc Inform Res, 2016, 22(3): 156-163. [doi: 10.4258/hir.2016.22.3.156]

[2] Zhou K, Fu C, Yang S. Big data driven smart energy management: From big data to big insights. Renewable & Sustainable Energy Reviews, 2016, 56: 215-225.[doi: 10.1016/j.rser.2015.11.050]

[3] Cerchiello P, Giudici P. Big data analysis for financial risk management. Journal of Big Data, 2016, 3(1): 18.

[4] Feng DG, Zhang M, Li H. Big Data Security and Privacy Protection. Chinese Journal of Computers, 2014, 37(1): 246-258. [doi: 10.3724/SP.J.1016.2014.00246]

[5] Sandhu R. The future of access control: Attributes, automation and adaptation. IEEE International Conference on Information Reuse and Integration, 2013: xxiii-xxiv.[doi: 10.1109/IRI.2013.6642437]

[6] Meng XF, Ci Xiang. Big Data Management: Concepts, Techniques and Challenges. Journal of Computer Research and Development, 2013, 50(1):146-169.

[7] Li H, Zhang M, Feng DG, Hui Z. Research on Access Control of Big Data. Chinese Journal of Computers, 2017, (1): 72-91.

[8] Liu AD, Du XH, Wang N, Li SZ. Survey on Information Security Techniques for Blockchain Technology. Journal of Software, 2018,29(7): 2092-2115.

[9] Yuan E, Tong J. Attributed based access control (ABAC) for Web services. IEEE International Conference on Web Services, 2005. ICWS 2005. Proceedings, 2005: 561-569. [doi: 10.1109/ICWS.2005.25]

[10] Fang L, Yin LH, Guo YC, Fang BX. A Survey of Key Technologies in Attribute-Based Access Control Scheme. CHINESE JOURNAL OF COMPUTERS, 2017, 40(7): 1680-1698.

[11] Hui Z, Li H, Zhang M, Feng DG. Risk-adaptive access control model for big data in healthcare. Journal on Communications, 2015, 36(12): 190-199.

[12] Ray I, Kumar M, Yu L. LRBAC: A Location-Aware Role-Based Access Control Model. International Conference on Information Systems Security, 2006: 147--161.[doi: 10.1007/11961635_10]

[13] Damiani M L, Bertino E, Catania B, et al. GEO-RBAC: A spatially aware RBAC. Acm Transactions on Information & System Security, 2007, 10(1): 2.[doi: 10.1145/1210263.1210265]

[14] Frank M, Buhman J M, Basin D. Role Mining with Probabilistic Models. ACM, 2013: 1-28.

[15] Molloy I, Chen H, Li T, Wang Q, Li N, Bertino E, Calo S, Lobo J. Mining roles with semantic meanings. SACMAT 2008, ACM Symposium on Access Control MODELS and Technologies, Estes Park, Co, Usa, June 11-13, 2008, Proceedings, 2008: 21-30.

[16] Vaidya J, Atluri V, Guo Q. The role mining problem. Acm Transactions on Information & System Security, 2010, 13(3): 1-31.

[17] Yang K, Jia X, Ren K. Secure and Verifiable Policy Update Outsourcing for Big Data Access Control in the Cloud. IEEE Transactions on Parallel & Distributed Systems, 2015, 26(12): 3461-3470.[doi: 10.1109/TPDS.2014.2380373]

[18] Yang K, Han Q, Li H, Zheng K, Su Z, Shen X. An Efficient and Fine-Grained Big Data Access Control Scheme With Privacy-Preserving Policy. IEEE Internet of Things Journal, 2017, 4(2): 563-571. [doi: 10.1109/JIOT.2016.2571718]

[19] Cao ZF, Dong XL, Zhou J, Shen JC, Ning JT, Gong JQ. Research Advances on Big Data Security and Privacy Preserving. Journal of Computer Research and Development, 2016, 53(10): 2137-2151. [doi: 10.7544/issn1000-1239.2016.20160684]

[20] Li FH, Su M, Shi GZ, Ma JF. Research Status and Development Trends of Access Control Model. Acta Electronica Sinica, 2012, 40(4): 805-813.

[21] Joshi J B D, Bhatti R, Bertino E, et al. Access-Control Language for Multidomain Environments. IEEE Internet Computing, 2004, 8(6): 40-50. [doi: 10.1109/MIC.2004.53]

[22] Lee H K, Luedemann H. lightweight decentralized authorization model for inter-domain collaborations. ACM Workshop on Secure Web Services, 2007: 83-89. [doi: 10.1145/1314418.1314431]

[23] Maesa D D F, Mori P, Ricci L. Blockchain Based Access Control. IFIP International Conference on Distributed Applications and Interoperable Systems, Springer, 2017: 206-220. [doi: 10.1007/978-3-319-59665-5_15]

[24] Zyskind G, Nathan O, Pentland A S. Decentralizing Privacy: Using Blockchain to Protect Personal Data. IEEE Security and Privacy Workshops, 2015: 180-184. [doi: 10.1109/SPW.2015.27]

[25] Ouaddah A, Abou Elkalam A, Ait Ouahman A. FairAccess: a new Blockchain-based access control framework for the Internet of Things. Security & Communication Networks, 2016, 9. [doi: 10.1002/sec.1748]

[26] Ouaddah A, Mousannif H, Elkalam A A, Ouahman A A. Access control in the Internet of Things: Big challenges and new opportunities. Computer Networks, 2017, 112:237-262. [doi: 10.1016/j.comnet.2016.11.007]

[27] Ouaddah A, Elkalam A A, Ouahman A A. Towards a Novel Privacy-Preserving Access Control Model Based on Blockchain Technology in IoT. Springer International Publishing, 2017. [doi: 10.1007/978-3-319-46568-5_53]

[28] Ouaddah A, Bouij-Pasquier I, Elkalam A A, Ouahman A A. Security analysis and proposal of new access control model in the Internet of Thing. International Conference on Electrical and Information Technologies, 2015: 30-35. [doi: 10.1109/EITech.2015.7162936]

[29] Azaria A, Ekblaw A, Vieira T, et al. MedRec: Using Blockchain for Medical Data Access and Permission Management. International Conference on Open and Big Data, 2016: 25-30. [doi: 10.1109/OBD.2016.11]

[30] Ekblaw A, Azaria A, Halamka J D, Md†, Lippman A. A Case Study for Blockchain in Healthcare: "MedRec" prototype for electronic health records and medical research data. Technical Report, 5-56-ONC, Massachu-setts Institute of Technology, 2016.

[31] Christidis K, Devetsikiotis M. Blockchains and Smart Contracts for the Internet of Things. IEEE Access, 2016, 4: 2292-2303.

[32] Tapscott, Alex, Tapscott. Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business and the World, 2016.

[33] Smart Contracts. http://szabo.best.vwh.net/smart.contracts.html.

[34] Vitalikbuterin. Ethereum white paper, 2013.

[35] Chen GK, Yin XL, Liu WL. Access Control Model Applicability for Big Data. Authentication and Confidentiality, 2016, 7(7): 3-5.

[36] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. Consulted, 2008.

[37] Ding W, Wang GC, Xu AD, Chen HJ, HONG C. Research on Key Technologies and Information Security Issues of Energy Blockchain. Proceedings of the CSEE, 2018, 38(4).

附中文参考文献:

[4] 冯登国, 张敏, 李昊. 大数据安全与隐私保护. 计算机学报, 2014, 37(1): 246-258.

[6] 孟小峰, 慈祥. 大数据管理:概念、技术与挑战. 计算机研究与发展, 2013, 50(1):146-169.

[7] 李昊, 张敏, 冯登国, 惠榛. 大数据访问控制研究. 计算机学报, 2017, (1): 72-91.

[8] 刘敖迪, 杜学绘, 王娜, 李少卓. 区块链技术及其在信息安全领域的研究进展. 软件学报, 2018,29(7): 2092-2115.

[10] 房梁, 殷丽华, 郭云川, 方滨兴. 基于属性的访问控制关键技术研究综述. 计算机学报, 2017, 40(7): 1680-1698.

[11] 惠榛, 李昊, 张敏, 冯登国. 面向医疗大数据的风险自适应的访问控制模型. 通信学报, 2015, 36(12): 190-199.

[19] 曹珍富, 董晓蕾, 周俊, 沈佳辰, 宁建廷, 巩俊卿. 大数据安全与隐私保护研究进展. 计算机研究与发展, 2016, 53(10): 2137-2151. [doi: 10.1109/JIOT.2016.2571718]

[20] 李凤华, 苏铓, 史国振, 马建峰. 访问控制模型研究进展及发展趋势. 电子学报, 2012, 40(4): 805-813.

[35] 陈垚坤, 尹香兰, 刘文丽. 大数据环境下访问控制模型适用性研究. 信息安全与技术, 2016, 7(7): 3-5.

[37] 丁伟, 王国成, 许爱东, 陈华军, 洪超. 能源区块链的关键技术及信息安全问题研究. 中国电机工程学报, 2018, 38(4).