可编辑区块链: 模型、技术与方法

袁 勇 [1, 2] 王飞跃 [1, 3]

摘 要 可编辑区块链是区块链领域新兴而颇有争议的热点课题, 致力于在保障区块链安全可信等良好性质的前提下实现链上数据的可控编辑操作. 本文系统性地梳理和研究了可编辑区块链技术在信息安全和监管等领域面临的现实需求, 提出了可编辑区块链的工作框架, 并从数据修改、删除、插入、过滤和隐藏五个环节详细阐述了可编辑区块链的技术与方法, 最后讨论了该领域亟需解决的若干关键问题.

关键词 区块链, 可编辑区块链, 比特币, 变色龙哈希函数

引用格式 袁勇, 王飞跃. 可编辑区块链: 模型、技术与方法. 自动化学报, 2020, **46**(5): 831−846

DOI 10.16383/j.aas.2020.y000002

Editable Blockchain: Models, Techniques and Methods

YUAN Yong[1, 2] WANG Fei-Yue[1, 3]

Abstract The editable blockchain is a novel but controversial topic in blockchain research, aiming at editing the blockchain data without undermining their desirable features such as security and trustability. In this paper, we systematically analyze and address the requirements for editable blockchains in information security and regulation, and present a working framework of editable blockchains. We also investigate the detailed models, techniques and methods in redacting, deleting, inserting, filtering and hiding data on blockchains, and discuss some open problems. Our purpose is to provide helpful reference and guidance for future research and development in this novel area.

Key words Blockchain, editable blockchain, Bitcoin, chameleon hash function

Citation Yuan Yong, Wang Fei-Yue. Editable blockchain: models, techniques and methods. *Acta Automatica Sinica*, 2020, **46**(5): 831−846

区块链是以比特币为代表的数字加密货币体系的核心支撑技术, 是一种全新的去中心化基础架构与分布式计算范式[1−2]. 区块链的技术特点可以归纳为 “TRUE” 和 “DAO”, 前者表示可信 (Trustable)、可靠 (Reliable)、可用 (Usable) 和高效 (Effective, efficient), 其中高效指的是区块链技术将带来社会运行效率和效果的提高; 而后者则表示分布式与去中心化 (Distributed, decentralized)、自主性与自动化 (Autonomous, automated)、组织化和有序性 (Organized, ordered)[3]. 因此, 区块链被视为引发下一次产业革命的核心要素之一, 有助于打造未来的智能产业和数字经济新生态. 2019 年 10 月, 中央政治局第十八次集体学习聚焦区块链技术, 强调 “区块链是我国核心技术自主创新的重要突破口”. 国际如 Facebook、IBM、摩根大通, 国内如百度、腾讯、阿里巴巴等企业相继布局区块链技术. 显然, 区块链已经站在新一代信息技术的最前沿.

区块链技术自诞生伊始就带有极其鲜明的技术特色. 其中, 去中心化和不可篡改无疑最具革命性, 被认为是区块链机器的 “信任之源”. 正是基于这些技术特色, 区块链可以使得互不信任的分布式节点 “共享” 同一份数据账本、通过共识算法和激励机制实现区块链的协同 “共治” 、最终 “共建” 一个安全可信的生态系统, 并可望在数字经济和社会治理等领域提供重要的技术支撑[4]. 然而, 近年来的应用实践使得人们逐渐意识到: 去中心化和不可篡改性是一柄双刃剑, 其在为区块链数据奠定坚实的安全和信任基础的同时, 也极大地限制了区块链技术在实践中的应用前景.

因此, 区块链相关研究和应用实践正呈现出从 “乌托邦” 回归现实的趋势. 一方面, 完全去中心化的公有链 (非授权区块链) 受性能、效率和部署成本

收稿日期 2020-01-28 录用日期 2020-04-01
Manuscript received January 28, 2020; accepted April 1, 2020
国家重点研发计划 (2018AAA0101401), 国家自然科学基金 (61533019, 71702182) 资助
Supported by National Key R&D Program of China (2018AAA0101401), National Natural Science Foundation of China (61533019, 71702182)
本文责任编委 刘德荣
Recommended by Associate Editor LIU De-Rong
1. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室 北京 100190 2. 青岛智能产业技术研究院平行区块链技术创新研究中心 青岛 266109 3. 中国科学院哲学研究所 北京 100000
1. The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190 2. Innovation Center for Parallel Blockchain, Qingdao Academy of Intelligent Industries, Qingdao 266109 3. Institute of Philosophy, Chinese Academy of Sciences, Beijing 100000

等限制, 短期内难以在某些关键业务场景和领域得到广泛应用, 因而催生了服务于现实场景的、自主可控的 “主权” 区块链 (授权区块链), 例如多中心化的联盟链和完全中心化的私有链. 记账权重新回归到少数人手中[5−6]. 另一方面, 不可篡改性使得数据一旦上链即可安全和永久存储, 然而在实践应用中, 这种特性使得区块链为各类虚假新闻和不良信息提供了更好的发布和传播渠道, 对区块链信息内容安全乃至社会舆论环境带来负面影响, 因而迫切需要安全、便捷、可控的技术手段来更新链上关键数据并清除有害数据[7].

本文主要关注区块链数据编辑技术. 就研究现状而言, 该领域目前虽是小范围探索, 但有实际应用和国家监管的双重需求驱动, 因而在最近两三年内快速发展. 现有文献中, 支持数据编辑操作的区块链研究通常冠以可编辑 (Editable)、可修改 (Redactable/correctable)、可重写 (Rewritable) 或者可变异 (Mutable) 等字样, 其含义略有差异. 在本文中, 我们认为 “可编辑” 的研究范畴更大, 并以可编辑区块链统称那些支持针对链上数据的增、删、改等操作的区块链.

本文组织结构如下: 第 1 节主要介绍可编辑区块链的现实需求和研究框架; 第 2 节给出了区块链的模型表示以及变色龙哈希函数、秘密共享等关键技术; 第 3 节至第 7 节从数据修改、删除、插入、过滤和隐藏五个方面, 系统性地阐述了可编辑区块链的模型、技术与方法; 第 8 节讨论了可编辑区块链亟需解决的问题与挑战, 以及未来研究方向; 第 9 节总结全文.

1 可编辑区块链: 现实需求与研究框架

1.1 现实需求

可编辑区块链是近年来新兴而颇有争议的研究课题. 部分研究者认为不可篡改性是区块链技术不可动摇的重要根基, 而数据编辑技术将会在去中心化的记账权之上增加中心化的编辑权这一 “漏洞”, 从而使得区块链数据在达成共识并上链后, 再次面临着中心化、甚至是恶意的数据篡改. 这种观点固然存在其内在的合理性, 然而从区块链应用实践来看, 目前区块链技术在信息监管、隐私保护、数据更新、可扩展性等四个方面都存在切实的数据编辑需求, 迫切需要研究和应用可编辑区块链技术.

1) 从信息监管角度来讲, 目前区块链领域的研究和应用更多地强调链上数据的存储与传输安全, 而忽略了更为重要的信息内容安全, 主要表现在: a) 缺乏必要的上链信息审核与评估机制. 大多数区块链系统的验证者 (矿工) 重点核查上链信息的语法正确性, 而忽略了信息内容的语义合理性甚至是真实性[8]; b) 链上数据的搜索与甄别机制尚不完善, 传统的互联网大数据分析、监控与预警技术短期内难以应用于新兴的区块链架构, 因此不良信息上链后很难及时、准确地发现与预警; c) 区块链的公开透明性和不可篡改性, 使得不良信息将带来大范围、持续性甚至是永久性的负面影响. 由此可见, 区块链已成为规避监管、发布不良信息的有效途径, 任何人都可以极低的成本发布信息, 从而使得区块链处于舆论监管的 “失控” 状态, 为国家信息安全和网络安全带来潜在威胁. 另外, 比特币和以太坊等区块链已被证实存在色情链接、计算机病毒、僵尸网络等内容, 从区块链用户角度来讲, 加入并运行区块链节点就意味着存储和传播这些非法内容, 从而将面临着潜在的法律风险.

2) 从隐私保护角度来讲, 区块链技术可能与某些强调保护用户隐私、规避敏感内容的法律法规相悖. 例如, 2018 年欧盟出台的《通用数据保护条例 (General Data Protection Regulation, GDPR)》中明确规定: 用户具有 “被遗忘权 (Right to be forgotten), 即用户个人可以要求责任方隐藏或者删除关于自己的隐私数据记录, 这对于既无中心化责任方又不可篡改的区块链来说显然是不可实现的[9]. 虽然近年来以隐私保护为目的的加密货币 (如 ZCash 和 Monero) 快速发展, 然而要从根本上实现区块链的 GDPR 兼容性, 就必须首先实现链上数据的可编辑性.

3) 从数据更新角度来讲, 区块链的技术本质之一是去中心化的数据库, 其在数据采集、传输、验证直至上链过程中都可能会存在由于主观故意或者客观疏忽而导致的错误数据, 因而需要数据编辑技术来修改错误数据、更新陈旧数据. 目前处理这类错误数据的方式主要是硬分叉, 例如 2016 年以太坊 “The DAO” 项目由于智能合约的漏洞而导致的社区分裂和链上硬分叉. 这对于区块链生态来说是巨大的安全隐患和资源浪费[10−11].

4) 从可扩展性角度来讲, 区块链 (特别是公有链) 数据规模的不断增长, 使得存储和验证链上数据的开销不断增加. 数据显示, 截止到 2020 年初, 比特币全节点的区块链账本已经超过 200 GB 且每年稳定地增长约 52 GB, 以太坊全节点账本已经超过 400 GB, 存档节点数据规模已经超过 4 TB. 一方面, 持续增长的全节点存储空间将导致全节点比例持续下降, 从而使得区块链中心化趋势加剧. 另一方面, 区块链节点验证历史数据的计算开销也必将提高[12]. 因此, 在不破坏区块链完整性和安全性的前提下, 适当地删除非关键历史数据将是提高区

块链性能和可扩展性的重要手段.

由此可见, 可编辑区块链具有明确而迫切的现实需求. 通过技术创新和机制设计, 实现区块链的可编辑性和安全可信性的有机融合, 将能够进一步促进区块链技术脱虚向实、实现大规模落地应用.

1.2　研究框架

本文提出的可编辑区块链研究框架如图 1 所示. 一般来说, 可编辑区块链通常从编辑类型、编辑对象、编辑模态、编辑架构和控制策略五个侧面加以研究.

1) 按照数据操作类型, 可以分为修改、删除、插入、过滤和隐藏共五类编辑操作. 其中, 数据过滤是面向上链前数据的筛选和净化过程, 致力于在数据上链之前, 最大限度地识别不良信息并阻止其通过区块链的共识验证过程; 其他四类则均是面向链上数据的操作. 理论上讲, 数据修改是普适性的技术, 即支持数据任意修改的区块链技术必然也支持数据的任意插入、删除和隐藏.

2) 按照数据编辑对象, 可以分为区块级、交易级和数据项级编辑操作. 区块级编辑技术粒度最大, 只可以替换完整的区块而无法精准定位和修改区块中的特定数据; 交易级和数据项级编辑技术粒度相对较小, 前者重点针对区块中的金融交易数据 (例如交易金额和接收方地址等), 而后者则侧重于非金融文本数据 (例如 OP_RETURN 类交易附言或其他文本数据). 通常来说, 交易级编辑技术将会改变区块链内部的交易逻辑流和价值分配体系, 是强上下文相关的编辑操作, 因而难度更高. 例如, 简单的修改一笔交易的金额可能就会凭空 “创造” 或者 “燃烧” 一定数量的加密货币、进而导致其后续交易失效.

3) 按照数据编辑模态, 可以分为中心化、多中心化和去中心化的数据编辑, 表示数据编辑权限 (包括请求权、验证权和修改权) 是否属于特定的中心化机构或者实体. 可编辑区块链中, 面向记账权竞争的共识过程与面向编辑权竞争的共识过程可能是相对独立的, 因此上述三种编辑模态并不一定与私有链、联盟链和公有链一一对应 (尽管大多数情况确实如此).

4) 按照数据编辑架构, 可以分为单链架构和平行链架构. 前者仍然维护单一的线性区块链条, 通过变色龙哈希函数等特定的技术手段实现区块数据的定点物理修改、或者在后续区块中追加修改后的新数据; 后者则是维护独立运行的两条或多条平行链实现数据修改, 例如两条平行的区块链或者两条平行的哈希链等.

5) 区块链编辑过程一般有明确的控制策略, 详细规定涉及的数据范围 (哪些数据可以被编辑)、编辑权限 (如何确定谁有请求编辑的权限、谁有验证新数据正确性的权限、以及谁有最终实施编辑操作的权限)、编辑流程 (实施编辑操作有哪些具体步骤)、约束规则 (编辑过程中涉及哪些规则与约束条件) 等要素. 控制策略的设计与实施一般取决于实际场景需求.

值得一提的是, 在分布式和去中心化的区块链系统中, 真正在所有节点上完全实现修改、删除等编辑操作是不可能实现的. 部分区块链节点可以通过单方面地不执行编辑操作、拒绝升级甚至硬分叉等手段来保存修改前的数据. 在这些情况下, 目前尚缺乏有效手段实现数据的强制编辑. 此外, 现有文献中还包括链上数据的搜索、查询、分析等操作的相关研究, 但这些操作并不会改变链上数据, 因此我们认为其不属于可编辑区块链的研究范畴.

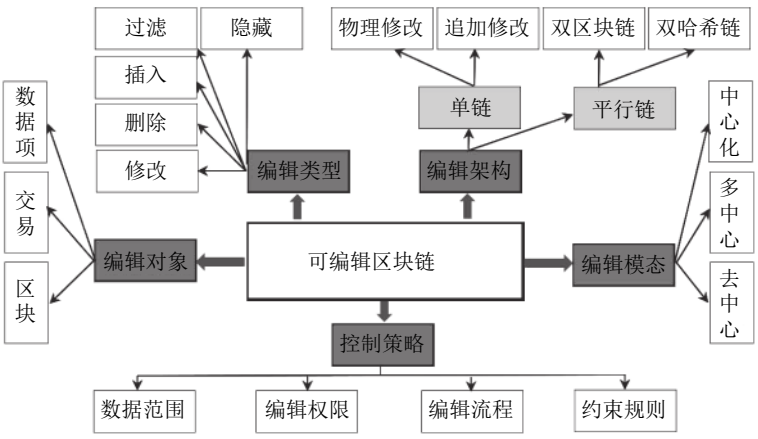

图 1　可编辑区块链的研究框架

Fig. 1　Research framework of editable blockchain

2 可编辑区块链: 模型与技术

2.1 模型与符号表示

借鉴现有文献 [13−14], 本文采用如下区块链模型和符号表示体系: 假设区块链 C_N 是由首尾相连的区块组成的线性链条, 其中 N 为最新区块的高度; 每个区块记为三元组 $B_i = \langle s_i, x_i, ctr_i\rangle$, $i \in [0, N]$, B_0 和 B_N 分别为创世区块和最新区块 (头部). 三元组中, $s_i \in \{0,1\}^K$ 是长度为 K 位的前一区块哈希值, $x_i \in \{0,1\}^*$是任意长度的当前区块数据, $ctr_i \in \mathbf{N}$ 是当前区块共识过程生成的随机数 Nonce. 如果下一区块记为 $B_{N+1} = \langle s_{N+1}, x_{N+1}, ctr_{N+1}\rangle$, 则其上链后形成的新区块链记为 $C_{N+1} = C_N||B_{N+1}$, 且区块间哈希链接 $s_{N+1} = H(ctr_N, G(s_N, x_N))$, 其中 $H: \{0,1\}^* \to \{0,1\}^K$ 和 $G: \{0,1\}^* \to \{0,1\}^K$ 是长度为 K 位的抗碰撞哈希函数, 分别称为外哈希函数和内哈希函数. 为保证区块链的有效性和完整性, 必须有如下两式成立:

$$\begin{aligned} Valid_q^D(B_i) &= (H(ctr_i, G(s_i, x_i)) < D) \cap (ctr_i < q) = 1 \\ s_{i+1} &= H(ctr_i, G(s_i, x_i)) \end{aligned} \tag{1}$$

其中参数 $D \in \mathbf{N}$ 是区块链当前难度, q 是每一轮共识过程中最大允许的哈希请求数. 此外, 令 $C_\emptyset$ 表示空链, $C^{\lceil k}$ 表示去掉链条上较新的 k 个区块后的区块链, 而 $C^{\lfloor k\rfloor}$ 则表示去掉历史上较旧的 k 个区块后的区块链.

2.2 变色龙哈希函数

基于哈希运算生成的区块之间的哈希链路是区块链数据极难篡改的重要原因. 常见的哈希函数具有抗碰撞性, 即任意寻找两组不同的数据 m 和 m', 使其哈希函数 $Hash(m) = Hash(m')$ 在计算上是不可行的; 同时, 哈希函数具有高灵敏度, 即使输入数据 m 发生一个比特位的微小修改, 输出哈希值 $Hash(m)$ 也会发生明显的改变. 因此, 如果将区块 B_i 的数据 x_i 修改为 x_i', 则必有 $s_{i+1} \neq Hash(ctr_i, G(s_i, x_i'))$, 从而破坏哈希链路的完整性. 因此, 现有研究的基本思路是采用变色龙哈希函数, 在不改变哈希函数输出结果的前提下实现区块数据的任意修改.

变色龙哈希函数是由 Krawczyk 和 Rabin 提出的一种带陷门的单向哈希函数[15−16]. 如果掌握陷门信息, 则可以轻易地计算任意输入数据的哈希碰撞, 从而可以在不改变哈希函数输出的情况下, 任意地改变哈希函数的输入. 如果不掌握陷门信息, 则变色龙哈希函数与传统的哈希函数一样具有抗碰撞性. 因此, 如果将区块链的哈希函数 (例如内哈希函数 G) 替换为变色龙哈希函数, 并且人为地设置陷门, 则可以任意修改区块数据而不会破坏哈希链路的完整性.

形式上, 变色龙哈希函数可以定义为: 对于任意数据 x 和随机选择的参数 r, 给定一个陷门 tk, 可以找到消息对 (x, r) 和 (x', r'), 使得 $CH(x, r) = CH(x', r')$, 此处 CH 为变色龙哈希函数. 在可编辑区块链应用中, x 和 x' 分别对应原区块数据和修改后的区块数据, 且 $x \neq x'$. 变色龙哈希函数通常有如下四个算法, 即密钥生成算法 HG、哈希生成算法 CH、哈希验证算法 HV 以及哈希碰撞算法 HC :

1) 密钥生成算法 $HG(1^n) = (hk, tk)$: 生成变色龙哈希的公钥 hk 和私钥 (陷门) tk, n 为安全性参数;

2) 哈希生成算法 $CH(hk, x; r) = (h, \xi)$: 给定公钥 hk、任意数据 x 和随机数 r, 生成哈希值 h 和随机数 ξ;

3) 哈希验证算法 $HV(hk, x, (h, \xi))$: 给定公钥 hk、任意数据 x、哈希值 h 和随机数 ξ, 如果 (h, ξ) 是正确的哈希值, 则输出 1, 否则输出 0;

4) 哈希碰撞算法 $HC(tk, (h, x, \xi), x')$: 给定陷门 tk、三元组 (h, x, ξ) 和数据 x', 输出新随机数 ξ', 使得 $HV(hk, x, (h, \xi)) = HV(hk, x', (h, \xi')) = 1$.

显然, 掌握陷门密钥就意味着拥有区块链的修改权, 因此陷门密钥的管理对于变色龙哈希函数来说至关重要. 对于私有链来说, 陷门密钥一般由可信的中心化验证者掌握, 可以实现任意链上数据的修改操作; 而对于多中心和去中心化的联盟链和公有链来说, 陷门密钥则须在多个 (固定的或者可变的) 验证者之间共享, 由验证者之间的共识过程决定是否和如何修改区块数据, 这种针对修改权的共识过程可以与验证区块数据的共识过程相互独立. 从安全角度考虑, 掌握陷门密钥和数据修改权的验证者必须是事前不可预测的, 以避免针对性的安全攻击. 因此, 现有研究的常见思路一方面是每个验证者都可以拥有陷门密钥, 但所有验证者共同采用分布式随机数生成协议生成一个随机数, 并据此选择对应的验证者来实施修改操作[17]; 另一方面是基于秘密共享方案, 使得具有修改权限的验证者共享陷门密钥, 并通过投票共识过程来决定和实施修改操作[14, 18−19].

2.3 秘密共享

秘密共享是由 Shamir 提出的密码学算法[20], 可以实现陷门密钥在验证者之间的安全共享. 例如, (k, n) 秘密共享方案的基本思想是将陷门密钥以适当的形式拆分成 n 个份额并分发给不同的验证者管

理, 单个或者少数验证者利用其份额无法恢复完整的陷门密钥, 只有不少于 k 个验证者共同协作, 才能得到完整的陷门密钥. 形式上, (k,n) 秘密共享方案由如下两个算法组成:

1) 秘密分发算法 $Share(tk)=(\tau_1,\tau_2,\cdots,\tau_n)$: 给定陷门密钥 tk, 随机生成 n 个份额 $(\tau_1,\tau_2,\cdots,\tau_n)$;

2) 秘密恢复算法 $Rec(\tau_1,\tau_2,\cdots,\tau_n)=tk\vee\perp$: 给定 n 个份额 $(\tau_1,\tau_2,\cdots,\tau_n)$, 恢复陷门密钥 tk 或者返回异常值 $\perp$.

利用秘密共享方案来管理陷门密钥, 可以使得多个验证者共同协作来管理链上数据的修改权限, 避免单个或者少量恶意验证者获得陷门后随意篡改数据, 从而提高区块链数据的安全性和可信度. 一般来说, k 越大, 则秘密共享方案的安全性越高, 但可靠性越低. 对于修改链上数据来说, 通常可以取 $k\in(n/2,n)$, 即超过半数验证者同意修改, 则可以执行修改操作.

3　数据修改技术

本节将阐述若干具有代表性的链上数据修改技术. 限于篇幅, 本节重点阐述其技术思路和特点, 略去相关的背景知识和详细的算法细节.

3.1　单链条物理修改模式

2017 年, Ateniese 等最早提出了基于变色龙哈希函数的区块链数据修改方案[14], 具有兼容性强、适合目前主流的区块链架构等优良特点, 特别适用于带有少数可信验证者的授权区块链. 该方案的思路是将区块链的内哈希函数 G 替换为变色龙哈希函数 CH, 即 $G(s_i,x_i)\to CH(hk,(s_i,x_i);r)$, 从而可以在不改变外哈希函数 H、不破坏哈希链路完整性的情况下, 实现“单一”区块链上数据的“物理”修改.

首先, 对于完全中心化的私有链来说, 假设存在唯一的中心化验证者 $Validator$ 并且掌握变色龙哈希的陷门密钥 tk. 对于区块链 C 上每个待修改的区块 B_i 来说, 将变色龙哈希值 (h_i,ξ_i) 增加到 B_i 中; $Validator$ 将基于新数据 $\{x_i'\}_{i\in I}$ 计算内哈希函数的哈希碰撞, 将旧数据 x_i 修改为 x_i' 并生成新区块 B_i' (步骤 $4\sim6$). 所有区块修改完毕后, $Validator$ 将新链 C' 广播至区块链系统中, 其他验证者必须同步至新链 C' (即使系统中存在未修改的更长链条).

具体步骤如算法 1 所示[14].

算法 1. 中心化区块链的变色龙哈希修改算法

输入: 区块链 C, 修改位置集合 $I\subseteq[0,N]$, 新数据 $\{x_i'\}_{i\in I}$, 变色龙哈希函数陷门密钥 tk.

输出: 修改后的新区块链 C'.

1: $C'\leftarrow C$

2: 提取 C' 的区块集合 $(B_0,\cdots,B_N)$.

3: 对任意 $i=0,1,\cdots,N$, 如果 $i\in I$, 则:

4:　提取区块 $B_i=\langle s_i,x_i,ctr_i,(h_i,\xi_i)\rangle$;

5:　$\xi_i'\leftarrow HC(tk,(h_i,s_i||x_i,\xi_i),(s_i||x_i'))$;

6:　$B_i'=\langle s_i,x_i',ctr_i,(h_i,\xi_i')\rangle$;

7:　$C'\leftarrow C'^{[N-i+1}||B_i'||C'^{i]}$

8: END

9: $Validator$ 广播新链 C'

10: 其他验证者同步为新链 C'

其次, 对于多中心或者去中心化区块链来说, 可以通过秘密共享技术将陷门密钥安全地分发给预定义的验证者集合. 该集合对于联盟链来说可以是固定的“联盟验证者”集合, 对于公有链来说既可以是全部验证者 (矿工) 集合, 也可以是按照特定规则选出的验证者集合 (例如算力排名前 10 位的矿工或者矿池). 需要修改链上数据时, 验证者将会启动安全多方计算协议, 以分布式方式共同执行算法 1, 从而实现链上数据修改. 需要说明的是, 由于采用了秘密共享和安全多方计算等较为复杂的密码学工具来管理变色龙哈希陷门密钥, 实验结果显示这种方案在验证者集合规模较大 (例如超过 200个) 时的性能比较差, 因此实际上并不适合完全去中心化的非授权公有链场景.

具体步骤如算法 2 所示.

算法 2. 去中心化区块链的变色龙哈希修改算法

输入: 区块链 C, 修改位置集合 $I\subseteq[0,N]$, 新数据 $\{x_i'\}_{i\in I}$, 验证者集合 V.

输出: 修改后的新区块链 C'.

1: 生成变色龙哈希函数密钥 $HG(1^n)=(hk,tk)$

2: 将私钥分发给验证者 $Share(tk)=(\tau_1,\tau_2,\cdots,\tau_n)$.

3: 提取 C 的区块集合 $(B_0,\cdots,B_N)$.

4: 每个 V 中的验证者 i 分布式地执行如下操作:

5:　若同意修改数据, 则发送 τ_i 给其他验证者

6:　若收到足够份额, 计算 $tk=Rec(\tau_1,\tau_2,\cdots,\tau_n)$

7:　执行算法 1 的步骤 $3\sim8$

8:　广播新链 C'

9: END

10: 其他验证者同步为新链 C'

Ateniese 方案是目前较为完备的数据修改方案, 可以实现单条区块链数据的物理修改, 并且已由咨询公司埃森哲 (Assenture) 获得授权专利. 然而, 该方案存在若干潜在的问题: 首先是只能进行区块级修改, 即必须完整地替换整个区块, 因而粒

度太大; 其次是缺乏内容验证, 算法修改数据时并不验证内容的正确性, 而仅通过变色龙哈希确保哈希链路的完整性, 因而掌握陷门的恶意用户可以任意篡改链上数据; 再次是存在陷门密钥曝光问题, 验证者可以通过哈希碰撞推断出陷门密钥, 而且一旦重构陷门后即可多次修改数据, 存在可控性和安全性隐患, 该问题有望通过临时陷门 (Ephemeral trapdoor) 来解决[21−22]; 最后是数据修改只需有权修改的验证者之间达成共识即可, 而不需要生成区块的验证者同意, 这不适合某些存在区块奖励、需要出块人同意的场景.

3.2 单链条追加修改模式

与 Ateniese 方案相比, Puddu 方案则提出可变交易 (Mutable transactions) 概念[23−24], 通过将区块链交易结构改进为可变结构, 使得用户和验证者可以在后续某个时刻以追加发布新交易的方式来扩展旧交易, 从而实现单链条、以交易为粒度的追加修改 (“软” 修改).

形式上, 可变交易定义为三元组 $\hat{T}=\langle\pi,\alpha,P\rangle$, 其中 $\pi=(\pi_1,\cdots,\pi_k)$ 是一组特定类型的交易集合, 其中只有一个交易被标识为活跃交易 (其标号记为 $\alpha,\alpha\in[1,k]$), 其他交易均为非活跃交易, P 为交易变更的控制策略; 任意交易 π_i 有如下三种可选类型:

1) 标准交易: 交易发送方发送给接收方并转移一定数量的加密货币, 其数据域可能携带小规模的任意数据; 类似于比特币普通交易.

2) 合约交易: 交易发送方发送给所有验证者, 合约代码包含在交易数据域中.

3) 空交易: 无发送和接收方, 数据域为空.

每一个可变交易都必须包含至少一个标准交易或者合约交易, 以及必须有一个空交易, 并且同一个交易集 π 中除空交易之外的所有交易都必须有相同的发送方和接收方. 发送方生成可变交易 $\hat{T}$ 之后, 必须指定一个缺省的活跃交易 π_α. 控制策略 P 指定了有权修改交易 $\hat{T}$ 的实体, 这个实体可以是交易的发送者或者接收者, 也可以是其他用户或者智能合约; 该实体将会在旧交易上链后的指定时间段 Δt (是由 P 指定的约束规则) 内增加新交易版本并扩展到交易集 π 中. Δt 期满后, 可变交易 $\hat{T}$ 将会被锁定, 不再允许修改.

可变交易在任一时刻有且仅有一个活跃交易. 当需要修改交易数据时, 可以通过发布特定的元交易 (Meta transaction) $T'=(Ref_{\hat{T}},\alpha')$ 来引用和扩展可变交易 $\hat{T}$ 的交易集合 π, 并指定新的活跃交易 $\pi_{\alpha'}\in\pi$ 来代替旧的活跃交易 π_α, 从而实现交易数据的修改. 区块链系统按照正常交易流程来处理元交易 T', 由用户或者智能合约生成并发布, 通过验证者的共识验证之后存储上链. 针对一个可变交易 $\hat{T}$ 可以同时发布多个元交易, 最后一个完成共识上链的元交易将会决定最终的活跃交易. 如果元交易中将活跃交易设置为空交易, 则该可变交易 $\hat{T}$ 将会被 (逻辑上) 删除. 由此可见, 通过不断地发送元交易来更新可变交易 $\hat{T}$ 的活跃交易, 就可以持续更新交易数据并且可以保留交易的历史版本.

元交易只能由控制策略 P 指定的合法发送者生成, 因此数据修改的权限取决于控制策略. 显然, 控制策略必须首先在验证者之间达成共识, 而这种共识一般独立于区块链数据验证的共识过程.

基于上述交易结构与交易逻辑的改进, 算法 3 和图 2 给出了区块链数据修改的简要步骤.

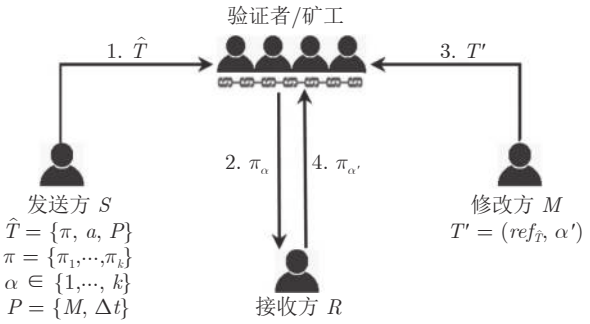

图 2 基于可变交易的区块链修改流程

Fig. 2 Redacting blockchain data based on mutable transactions

算法 3. 基于可变交易的区块链数据修改流程

输入: 区块链 C, 交易发送者 S 和接收者 R, 控制策略 $P=\langle M,\Delta t\rangle$, 验证者集合 V.

输出: 修改后的新区块链 C'.

1: S 发送可变交易 $\hat{T}$ 给接收者 R;

2: $\hat{T}$ 通过 V 共识验证后写入 C, 活跃交易为 π_α;

3: P 指定的修改者 M 在时间段 Δt 内生成指向 $\hat{T}$ 的元交易 $T'=(Ref_{\hat{T}},\alpha')$ 并发送至 V;

4: 若 T' 通过 V 的共识验证, 则将 $\hat{T}$ 的当前活跃交易 π_α 更新为 $\pi_{\alpha'}$ 并发送至 R.

由此可见, Puddu 方案实际上并不物理修改链上数据, 而是在后续区块中生成新的活跃交易, 旧的活跃交易并不会被删除或修改. 因此, Puddu 方案不会改变包含旧可变交易的区块数据, 因此也不会改变区块哈希和破坏哈希链路的完整性 (即不必使用变色龙哈希函数). 同时, 该方案保存了链上数据所有的历史版本和修改记录, 这可以支持区块链

数据的版本控制, 但另一方面却也难以有效清除或者隐藏链上的不良信息.

3.3　平行链模式

这种模式采用平行的两条链来实现数据修改. 就技术方案来讲, 可以分为双区块链模式和双哈希链两种实施模式.

3.3.1　双区块链模式

双区块链模式由 “共生” 的数据链和修正链组成, 其中数据链存储原始区块链数据, 而修正链则以区块为粒度存储修改后的新数据[25]. 验证者同时验证和维护两条区块链. 在顺序地遍历区块链时, 每次在数据链上发现区块被修改时, 就切换到修正链上读取相对应的新区块数据; 数据修改权限和修改内容是通过验证者之间基于共识的投票机制来共同决策, 因此这种模式特别适合去中心化的公有链.

数据修改流程和思路为: 数据链和修正链具有相同的创世区块, 如果某一时刻数据链的区块 B_{i+1} 的数据需要更新, 则请求此更新的用户可在数据链上发起一个选举交易 TX (Election transaction), 并提供待修改的区块高度、待修改的数据位置, 以及修改后的新区块 B'_{i+1} 等信息. 所有验证者根据控制策略 P 来验证 TX 中数据的正确性, 并在通过验证后将 TX 写入数据链上的最新区块中 (而非待修改区块中). 随后, 选举交易 TX 将会启动一次链上投票过程, 每个验证者将根据其是否同意此次数据修改进行投票 (以发送特定形式交易的方式), 这些投票将会陆续写入数据链的后续区块中. 投票持续时间 Δt (例如 1 000 个区块, 可由 P 定义) 截止后, 如果同意修改的验证者超过 P 中预定义的阈值比例 ρ, 则执行此次修改并将新区块 B'_{i+1} 写入修正链并重新编号. 否则数据修改请求将会被废弃. 上述修改流程的一个示例如图 3 所示.

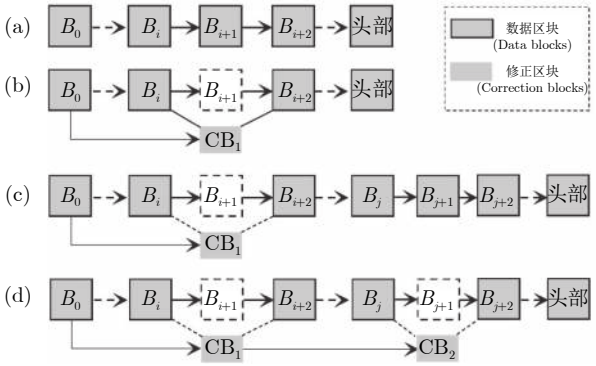

图 3　基于平行双区块链的数据修改流程

Fig. 3　Redacting blockchain data based on parallel blockchains

3.3.2　双哈希链模式

双哈希链模式仍然保留单条区块链, 但通过扩展区块结构、使得相邻两个区块之间保留两条哈希链接, 从而形成两条哈希链路[26]. 技术上, 这种模式在区块 $B_i = \langle s_i, x_i, ctr_i, y_i \rangle$ 中增加了一个 “Old state” 字段 y_i, 用以存储原始区块数据的 Merkle 根的一个副本, 其中 $y_i = G(s_i, x_i)$ 且区块生成后就不再变化. 当区块数据 x_i 修改为 x'_i 时, 新区块 $B'_i = \langle s_i, x'_i, ctr_i, y_i \rangle$ 的哈希值将会发生变化, 从而破坏了区块之间哈希链路的完整性; 然而由于其 Merkle 根的原副本使得 y_i 保持不变, 因此由 y_i 形成的第二条哈希链将仍然成立.

数据修改流程和思路如图 4 所示. 当区块 B_i 需要修改并替换为 B^*_i 时, 请求更新的用户将发起验证者的共识投票过程, 该过程与双区块链模式的投票过程相似 (此处不再赘述). 若超过一定比例的验证者同意修改, 则区块 B_i 将被替换为 B^*_i, 原区块 B_i 将被移除. 如图 4 所示, 完成修改后, 由于区块 B^*_i 中的数据 x_i 发生变化, 因此虚线所示的哈希链路被破坏, 而由 y_i 形成的实线哈希链路仍然成立, 从而保证了单条区块链的哈希完整性. 由此可知, 在数据共识过程中, 区块数据将按照正常流程验证, 若出现相邻区块的虚线链接失效, 则检查第二条使用 Old state 字段的实线链接是否有效, 如果有效则表示该区块曾被修改过, 否则表示该区块发生了未经授权的篡改.

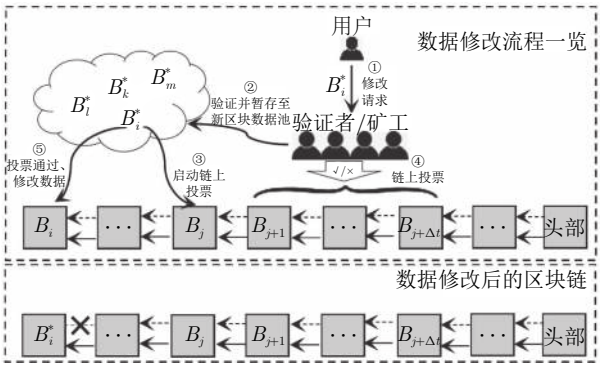

图 4　基于平行双哈希链的数据修改流程

Fig. 4　Redacting blockchain data based on parallel hash chains

综上所述, 两种平行链模式均适合非授权区块链, 不需要任何信任假设, 易于集成到比特币和以太坊等加密货币型区块链[27−28]. 链上除原有的数据共识之外, 同时针对控制策略 P、修改权限和修改内容也有独立的共识投票过程. 数据修改过程具有非常好的可解释性, 任何修改都可以被其他用户检

测和验证.

4 数据删除技术

链上数据删除技术主要包括本地数据删除和全局数据删除两种类型, 前者是指分布式节点可以独立地删除其本地部分数据, 以解决持续增长的区块链数据规模导致的存储瓶颈问题; 而后者则是指分布式节点通过共识算法来共同删除某些链上数据, 主要致力于解决清除链上不良信息的问题. 实际上, 由于区块链的分布式高冗余存储的特性, 一旦数据上链后, 目前尚无有效手段确保能够在所有节点上清除该数据, 因为恶意节点可以通过硬分叉等手段在其本地存储和检索不良信息, 即使这种不良信息已经在区块链全局视图中被删除了.

本节将详述区块链节点的本地数据删除和全局数据删除两种技术.

4.1 本地数据删除

近年来, 区块链数据存储规模随着交易数量和区块数量呈现出线性增长趋势. 由于新节点加入区块链系统时需要下载和处理全部数据以支持其独立地验证新交易, 因此具备全节点计算和存储能力的分布式节点越来越少, 使得区块链系统中心化趋势不断加剧[29−31].

为解决此问题, 中本聪在比特币创世论文中已经提出降低区块链账本规模的解决方案, 即回收磁盘空间 (Reclaiming disk space, RDS) 和简化支付验证 (Simplified payment verification, SPV)[1]. RDS 的核心思想是当比特币的最近一笔交易已经得到足够多的区块确认之后, 在其之前的历史交易就可以被删除以腾出磁盘空间, 但是区块的哈希值和 Merkle 树根必须保留以便验证区块或者交易. SPV 技术则允许轻节点不必存储全部数据, 当其验证交易信息时, 可以通过区块链网络向其他全节点发起查询请求获得所需数据.

RDS 和 SPV 之后, 现有文献中针对本地数据删除的研究一般称为区块链剪枝 (Blockchain pruning) 技术, 致力于使验证者在其本地区块链账本中删除单个交易或数据项、或者是删除特定时间点之前的全部历史数据[32−33].

2018 年, Palm 等提出了选择性交易剪枝技术, 即根据区块的状态可达性选择特定的不重要交易并删除[32]. 在工作量证明 (Proof-of-work, PoW) 等概率性共识系统中, 这种剪枝技术需要首先保证最近 n 个区块不被删除, 其中安全性参数 n 必须足够大以避免出现涉及 $m > n$ 个区块的重组情况, 这里 m 是已知的区块链重组中替换掉的最大区块数量. 换言之, 区块链 $C_N = \langle B_0, B_1, \cdots, B_N\rangle$ 中, $\langle B_{N-n+1}, \cdots, B_N\rangle$ 称为守护区块 (Guard block) 且不可删除, 其他区块的数据则可视交易状态的可达性来选择性地删除. 对于实用拜占庭容错 (Practical Byzantine fault tolerance, PBFT) 等确定性共识来说, 无须守护区块, 即 $n = 0$. 这里的 “状态” 是与区块数据相关联的数据结构, 可以通过遍历当前区块及其所有祖先区块的数据和交易而获得, 例如所有账户的余额就是典型的状态数据. 对于某一区块来说, 其现有的状态称为 “已达 (Derived)” 状态, 非现有但验证者有足够信息构建出的状态称为 “可达 (Derivable)” 状态, 非现有也无足够信息、必须从其他验证者处获得信息才能构建出的状态称为 “可获取 (Retrievable)” 状态.

根据状态可达性, 区块链中的交易可分为如下三类, 其中第二和第三类交易均可以删除:

1) 重要交易 (Significant): 即剪枝后会影响当前区块和后续区块的状态的交易. 例如图 5 中, 交易数据 $t_{1,1}$ 是重要交易, 因为删除 $t_{1,1}$ 之后, $a:1$ 和 $a:11$ 将不再属于状态 S'_1 和 S'_2.

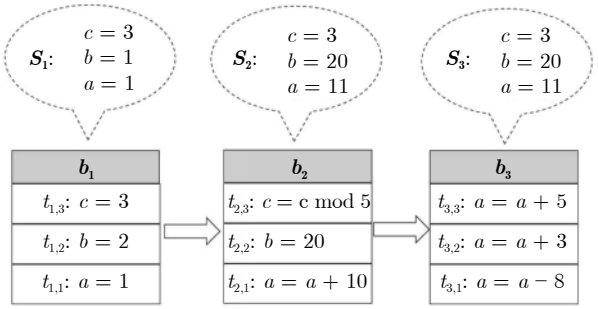

图 5 区块链交易的状态可达性示例

Fig. 5 An illustrative example of state derivability for blockchain transactions

2) 一般性非重要交易 (Universally insignificant): 如果在同一区块中删除一个或一组交易不会导致与该区块相关的状态变化, 则该交易称为一般性非重要交易. 例如图 5 中的 $t_{2,3}$ 被删除后不会导致状态 S'_2 发生变化, 再如交易集合 $b_3 = \langle t_{3,1}, t_{3,2}, t_{3,3}\rangle$ 被删除后也不会影响 S'_2 和 S'_3, 因此这些交易是一般性非重要交易;

3) 追溯性非重要交易 (Retroactively insignificant): 假定存在一组不再需要可达的状态集合 R, 如果删除一个或者一组交易只会影响 R 中状态的可达性, 而不会影响 R 之外的状态, 则称该交易为追溯性非重要交易; 例如图 5 中, 如果删除交易 $t_{1,2}$,

则状态 S_1' 不再可达、而 S_2' 不变. 如果 $S_1' \in R$, 则 $t_{1,2}$ 即为追溯性非重要交易.

在此基础上, 选择性交易剪枝技术的基本流程包括三个简单的步骤, 即 1) 准备阶段 (Preparation): 识别需要删除的区块并收集所需数据; 2) 标记阶段 (Marking): 检测交易的状态可达性和重要性, 标记所有非重要交易的位置; 3) 清除阶段 (Sweeping): 删除已标记的非重要交易. 具体算法细节详见文献 [32]. 实验结果显示, 在基于 Hyperledger fabric 的供应链场景用例中, 选择性交易剪枝技术可以使得区块链账本规模减少 84.49 %.

2019 年, Florian 等也提出了针对本地数据删除技术的剪枝方法[34]. 该方法致力于删除区块链上的单个数据块, 而非特定时间点之前的历史数据, 可以用来删除未花费交易输出 (Unspent transaction output, UTXO) 型区块链中存储于交易输出中的数据. 其他相似研究还包括 Mini-blockchain 方案[35]、Rollerchain 方案[33] 以及基于 PBFT 共识和分片的区块链剪枝方案[36] 等.

综上所述, 区块链剪枝技术是目前非常活跃的研究领域. 区块链剪枝可以部分地满足可扩展性和隐私保护要求, 但研究者认为剪枝技术将在一定程度上牺牲安全性 (即使在保留区块头的情况下). 同时, 本地剪枝将会导致区块链系统的 “公共地悲剧” 问题: 由于数据剪枝是分布式节点的个体行为, 而非群体共识行为. 因此, 一个 “理性” 的节点总是有动机执行本地数据剪枝过程, 即不存储全部数据, 而只存储必要数据 (例如最近 n 个区块的数据); 如果需要验证区块和交易, 则向网络中其他节点请求数据. 这种 “理性” 的结果会使得所有节点都不愿存储全部数据、以便节省计算和存储资源, 进而导致区块链的历史数据丢失. 由于缺乏历史数据, 新节点只能被迫下载最近的区块数据而且不得不信任该数据, 从而使得区块链账本失去信任基础.

4.2　全局数据删除

与本地数据的独立剪枝相比, 全局数据删除必须通过验证者的共识过程才能实现. 实际上, 现有文献中鲜见相关研究, 而是将全局数据的删除操作视为一种修改操作, 采用第 3 节所述的链上数据修改技术来实现. 如果需要删除整个区块, 例如在区块链 $C_N = \langle B_0, B_1, \cdots, B_N \rangle$ 中删除多个相邻区块 $\langle B_i, \cdots, B_{i+n} \rangle$, 则可以简单地将区块 $B_{i+n+1} = \langle s_{i+n+1}, x_{i+n+1}, ctr_{i+n+1} \rangle$ 中的哈希链接修改为 $s_{i+n+1} = H(ctr_{i-1}, G(s_{i-1}, x_{i-1}))$即可. 限于篇幅, 此处不再赘述.

5　数据插入技术

现阶段, 区块链数据插入技术的重点和难点是如何在去中心化的公有链中插入任意类型的数据. 公有链通常有大规模分布式节点的共识算力的维护, 因而安全性较高, 数据一旦上链后极难篡改, 这使得公有链成为永久保存重要数据的必然选择, 同时也为有害数据上链、规避国家信息监管提供了便利条件. 然而, 与数据类型灵活、链上数据操作可控性强的私有链和联盟链相比, 以比特币和以太坊等为代表的公有链主要存储以加密货币交易为核心的金融数据, 其数据结构和语法相对固定、数据上下文相关性强、数据插入空间和规模有限, 这些特点为插入任意类型的数据带来较高的难度[1, 37].

现有文献大多以比特币为原型来研究面向公有链的数据插入技术, 其他加密货币的技术原理相近, 只是在具体的数据插入位置和字段各有不同. 就比特币系统而言, 目前通常有四类数据插入手段, 分别将任意数据插入到 Coinbase 交易、OP_RETURN 脚本、P2X 类型交易脚本、以及非标准交易脚本 (极少数情况) 中[38–39], 如图 6 所示.

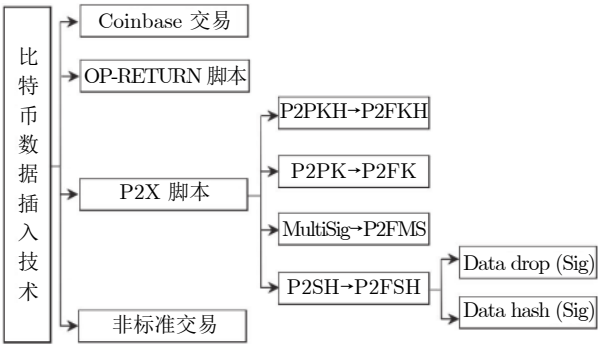

图 6　比特币数据插入技术框架

Fig.6　The framework of inserting data on Bitcoin blockchain

5.1　Coinbase 交易

每个比特币区块的第一个交易称为币基 (Coinbase) 交易. 该交易没有输入地址, 仅有一个输出地址, 其作用是将系统新生成的比特币奖励给成功 “挖出” 当前区块的矿工. Coinbase 交易没有解锁脚本 (即 ScriptSig 字段), 取而代之的是可变长度的 (最小 2 字节、最大 100 字节)、由矿工自定义的 Coinbase 数据. 因此, “挖出” 该区块的矿工可以在 Coinbase 交易中插入自定义的任意数据. 例如, 中本聪在比特币创世区块中插入的数据 “The Times 03/Jan/2009 Chancellor on Brink of Second Bail-

out for Banks (2009 年 1 月 3 日, 财政大臣正处于实施第二轮银行紧急援助的边缘)" 就是位于创世区块 Coinbase 交易的输入脚本中.

这种方式在实践过程中并不常用, 其原因首先是只有 "挖出" 当前区块的矿工才能够插入数据, 且最多只能插入 100 字节的任意数据, 因而成本极高且一般比特币用户无法使用; 其次是 Coinbase 数据会随着比特币协议的版本升级被赋予特定的含义, 使其不再适合插入任意数据.

5.2 OP_RETURN 脚本

OP_RETURN 是比特币交易的脚本操作码, 是专门设计用来存储额外的非金融数据 (Non-financial data) 的字段, 其作用相当于转账交易的备注或者附言信息. OP_RETURN 存在于比特币交易输出脚本中, 最多时可以在 ⟨*Data*⟩ 中存储 80 字节的任意数据 (字节数会随着比特币协议更新而波动)[40]. OP_RETURN 脚本规范如表 1 中 a) 所示.

OP_RETURN 脚本特别适合在比特币区块链上插入少量非金融数据. 该脚本具有上下文无关的特性, 在其中插入任意数据不会影响比特币系统内部交易流的逻辑自洽性, 而且比特币用户和矿工均可以通过创建新交易实现方便快捷的数据插入操作. 因此, 近年来该脚本的月均使用次数呈现出快速增长的趋势, 越来越多的用户和去中心化应用 (Decentralized application, DAPP) 选择使用 OP_RETURN 脚本来将外部数据 "锚定" 到比特币区块链, 使之永久存储、不可篡改[41]. 然而, OP_RETURN 脚本不适合通过高频生成新交易来插入大量数据, 因为这些新交易将通过分布式节点的概率性共识算法来实现排序, 其上链顺序具有较高的不确定性.

5.3 P2X 脚本

P2X (Pay to X) 脚本是比特币系统的 P2PKH (Pay to public key Hash)、P2PK (Pay to public key)、MultiSig (多重签名) 和 P2SH (Pay to script Hash) 等标准交易脚本的统称, 是将交易输出中锁定的比特币发送给由公钥、公钥哈希或者脚本等形式表示的接收者[42]. 由于比特币区块链具有伪匿名性, 分布式节点在共识过程中实际上并不验证当前交易的 "接收者" 是否真实存在, 而是在下一交易引用当前交易的 UTXO 时才会发现该交易是 "不可花费的", 即所谓的 "接收者" 无法花费该 UTXO. 因此, 通过策略性地操纵和修改 "接收者" 的标识, 即可以轻易地插入任意内容. 具体来说, P2X 方式是通过修改比特币交易输入 (解锁脚本)

表 1 比特币脚本规范与数据插入位置

Table 1 Scripts and locations of data insertion on bitcoin

序号	脚本名称	脚本规范	插入位置
a)	OP_RETURN	【输出脚本】OP_RETURN ⟨ Data ⟩	⟨ Data ⟩
b)	P2PKH	【输出脚本】OP_DUP OP_HASH160 ⟨ PubKeyHash ⟩ OP_EQUALVERIFY OP_CHECKSIG 【输入脚本】⟨ Sig ⟩ ⟨ PubKey ⟩	⟨ PubKeyHash ⟩
c)	P2PK	【输出脚本】⟨ PubKey ⟩ OP_CHECKSIG 【输入脚本】⟨ Sig ⟩	⟨ PubKey ⟩
d)	MultiSig	【输出脚本】M ⟨ PubKey 1 ⟩ ··· ⟨ PubKey N ⟩ N OP_CHECKMULTISIG 【输入脚本】OP_0 ⟨ Sig 1 ⟩ ··· ⟨ Sig M ⟩	⟨ PubKey ⟩
e)	P2SH	【输出脚本】OP_HASH160 ⟨ RedeemScriptHash ⟩ OP_EQUAL 【输入脚本】⟨ Data ⟩ ⟨ RedeemScript ⟩	⟨ Data ⟩ 或 ⟨ RedeemScript ⟩
	P2SH Data Drop	【输入脚本】⟨ Data ⟩ ⟨ Data ⟩ ⟨ Data ⟩ ⟨ Data ⟩ ⟨ RedeemScript ⟩ 【赎回脚本】OP_2DROP OP_2DROP ⟨ RandomNumber ⟩	
	P2SH- Data Drop/Sig	【输入脚本】⟨ Sig ⟩ ⟨ Data ⟩ ⟨ Data ⟩ ⟨ Data ⟩ ⟨ RedeemScript ⟩ 【赎回脚本】OP_DROP OP_2DROP ⟨ PubKey ⟩ OP_CHECKSIG	
	P2SH- Data Hash	【输入脚本】⟨ Data 1 ⟩⟨ Data 2 ⟩⟨ Data 3 ⟩⟨ RedeemScript ⟩ 【赎回脚本】OP_HASH160 ⟨ Data3Hash ⟩ OP_EQUALVERIFY OP_HASH160 ⟨ Data2Hash ⟩ OP_EQUALVERIFY OP_HASH160 ⟨ Data1Hash ⟩ OP_EQUAL	
	P2SH- Data Hash/Sig	【输入脚本】⟨ Sig ⟩⟨ Data 1 ⟩⟨ Data 2 ⟩⟨ Data 3 ⟩⟨ RedeemScript ⟩ 【赎回脚本】OP_HASH160 ⟨ Data3Hash ⟩ OP_EQUALVERIFY OP_HASH160 ⟨ Data2Hash ⟩ OP_EQUALVERIFY OP_HASH160 ⟨ Data1Hash ⟩ OP_EQUALVERIFY ⟨ PubKey ⟩ OP_CHECKSIG	

或输出 (锁定脚本) 中的公钥、公钥哈希、赎回脚本或其他脚本数据来插入任意数据.

5.3.1 P2FKH 方式

P2PKH 是最常见的比特币标准交易之一, 大多数比特币交易都是采用这种方式. 比特币脚本是基于逆波兰表示法的基于堆栈的执行语言, 因此 P2PKH 脚本的执行过程是通过顺序地执行输入脚本和输出脚本来检查公钥哈希的正确性以及数字签名的正确性. P2FKH (Pay to fake key Hash) 方式将任意数据存储到 P2PKH 输出脚本的公钥哈希 ⟨PubKeyHash⟩ 字段中, 即采用任意数据的伪公钥哈希代替原来正确的公钥哈希, 如表 1 中 b) 所示. 由于输入脚本中不可能存在正确的公钥和数字签名与伪公钥哈希相匹配, 因此该交易输出 UTXO 中 (通常是极少量) 的比特币将会被永久锁定、不可花费. 通常来说, 每笔比特币交易可以生成多个交易输出, 每个输出脚本中的公钥哈希 ⟨PubKeyHash⟩ 字段可以插入 20 字节的任意数据.

P2FKH 方式目前已被用来在比特币区块链中插入文本、图片和 MP3 文件. 这种方式虽然会永久锁定一部分比特币, 但其产生的 UTXO 是有效的, 矿工在共识阶段无法验证输出脚本中公钥哈希的正确性, 因而该 UTXO 将会永久存在于比特币的 UTXO 集合中, 造成 UTXO 集的 “膨胀”. 此外, P2FKH 方式每次插入数据都会 “燃烧掉” 少量比特币, 因此其成本较高.

5.3.2 P2FK 方式

P2FK(Pay to fake key) 方式与 P2FKH 方式的原理相似, 区别是将任意数据插入到 P2PK 输出脚本的公钥 ⟨PubKey⟩ 字段, 如表 1 中 c) 所示. 比特币公钥为 65 字节, 且 P2PK 的 OP 操作码较少, 因此效率要比 P2FK 稍高. 然而, 由于矿工可以较为容易地检测伪公钥, 因此这种方式并不常用. 显然, P2FK 方式同样存在永久锁定比特币和 UTXO 膨胀的问题, 此处不再赘述.

5.3.3 P2FMS 方式

MultiSig (多重签名) 也是比特币系统常见的标准交易形式. 一般来说, (M of N)形式的多重签名需要 N 个私钥中的至少 M 个才能授权支付. 顾名思义, P2FMS (Pay to fake MultiSig) 方式是将任意数据插入到交易输出脚本的 N 个公钥中, 如表 1 中 d) 所示. 例如 “1-of-3” 型多重签名地址可以将任意数据插入到 3 个公钥中, 每个未压缩的公钥为 65 字节 (总共 195 字节), 这会使得整个交易中的比特币被永久锁定, 因为没有对应的私钥可以与 3 个伪公钥匹配. 因此, 通常也可以将任意数据插入到其中 2 个公钥, 而保留一个真实公钥, 从而使得整个交易 UTXO 可以被唯一的真实私钥解锁. 每笔比特币交易可以有多个 P2FMS 输出, 因此可以利用这种方式存储一定数量的任意数据.

5.3.4 P2SH 方式

P2SH (Pay to ScriptHash) 是比特币改进提案 BIP 16 中提出的、可支持复杂交易以增强比特币可编程货币特性的标准交易脚本. P2SH 脚本将比特币发送到接收者定义的赎回脚本 (RedeemScript) 的哈希值所对应的地址. 因此, 与上述三种脚本是由发送者设置锁定条件不同, P2SH 脚本可以支持接收者以赎回脚本的形式灵活地设置转出条件. P2FSH (Pay to fake ScriptHash) 方式就是将任意数据插入到 P2SH 交易输出脚本中的赎回脚本哈希值 ⟨RedeemScriptHash⟩ 字段, 如表 1 中 e) 所示. 显然, 由于赎回脚本的哈希值被篡改, 接收方提供的赎回脚本将无法解锁 P2SH 脚本, 使得交易中的比特币将被永久锁定.

P2SH 交易的输入脚本也可以插入任意数据. 例如, Data Drop 方式是将任意数据插入到 P2SH 输入脚本中的数据 ⟨Data⟩ 字段或者赎回脚本自身 ⟨RedeemScript⟩(而非其哈希值), Data Hash 方式同样也是插入到 P2SH 输入脚本的 ⟨Data 1⟩···⟨Data 3⟩ 字段和赎回脚本 ⟨RedeemScript⟩ 中. 由于数据是插入到输入脚本, 因此 Data drop 和 Data Hash 方式生成的 UTXO 是可赎回和可花费的, 不会永久锁定比特币和造成 UTXO 集的膨胀.

5.4 非标准交易

极少数情况下, 通过将比特币标准交易脚本篡改为携带任意数据的非标准格式, 也可以实现数据上链. 虽然比特币区块链上确实存在这样的非标准交易, 但是使用这种方式插入数据很大程度上会被矿工在共识验证过程中过滤掉, 因而并不常见. 此处不再赘述.

综上, 比特币等公有链由于其极强的安全性和不可篡改性, 因而成为永久存储数据的重要手段. 根据不同的数据量和重要程度, 可以选择相对应的插入方式. 一般来说, OP_RETURN 和 Coinbase 交易适合插入少量数据, 而基于 P2SH 的 Data drop 和 Data Hash 方法则适合插入较多的数据. 需要说明的是, 事实上, 由于高冗余存储和高耗能共识等特性, 公有链并不适合存放大量非金融数据, 而是适合存放高价值密度、小规模的重要数据.

6 数据过滤技术

鉴于区块链的极难篡改特性, 数据过滤技术对

于避免虚假、敏感和有害等不良信息上链就显得尤为重要. 过滤技术就是在数据实际写入到区块链之前, 通过技术手段使得矿工在共识过程中有效地检测和识别不良信息, 或者通过经济手段提高不良信息上链的成本, 以达到过滤和净化上链数据的目的. 由于被过滤的数据并没有实际上链, 因此不会涉及链上数据修改, 也不会破坏区块之间哈希链路的完整性.

从技术角度来讲, 对于比特币等以金融交易为主的公有链来说, 通常无法严格地区分合法的哈希地址和任意二进制数据, 因此全节点不可能检测、识别和拒绝带有不良信息的交易; 对于联盟链和私有链来说, 目前针对上链数据通常是语法检查而非语义检查, 难以验证数据的真实性和可信性, 因而难以从根本上阻止不良信息上链; 由此可见, 目前尚无有效手段可以实现严格和可控的区块链数据过滤. 现有研究的基本思路主要是从提高不良信息上链的技术难度和经济成本两方面展开, 以求最大程度上过滤不良信息.

6.1 基于文本检测的数据过滤

这种方式假设数据量较大的交易中更有可能携带大量文本形式的非金融数据, 因此可以通过修改区块链的共识协议, 在共识算法的交易验证过程中增加文本检测器, 当交易中携带的文本数据量超过特定阈值时向矿工报警, 就可以阻止矿工打包这些交易、写入区块链.

例如, Matzutt 等提出采用自适应阈值方案设计文本检测器[43]: 如果交易的数据量较少 (例如 ≤ 100 B), 则允许上链: 对于中等规模数据量 (例如 ≤ 1 KB), 则认为可容忍但有风险; 对于较大规模数据量 (>1 KB), 则必须阻止其上链. 这里的阈值可自适应调整, 但在去中心化的公有链中通常需要矿工达成共识之后施行.

显然, 文本检测机制并不检查数据的语义, 因而无法从根本上防止小规模的不良信息上链; 同时也存在误报的风险, 即可能会阻止一些数据量较大的诚实交易上链.

6.2 基于经济成本的数据过滤

这种方式通过调节交易费来限制带有大量数据的交易, 从而提高任意数据上链的成本, 降低用户插入数据的动机.

例如, 给定交易 t 的数据规模为 z_t 且交易输出的数量为 n_t,则可设置该交易的强制性最小交易费为:

$$F(t)=\alpha\times(z_t+\beta(n_t)n_t) \tag{2}$$

其中 α 是按字节计费的基础费用, $\beta(n_t)$ 是区别小、中、大规模交易输出数量的函数, n_t 定义在阈值 T_s 和 T_m 上, 即交易输出的数量位于区间 $(0,T_s)$ 则为小交易, $[T_s,T_m]$ 为中等规模交易, 而 (T_m,∞) 为大交易. 例如, $\beta(n_t)$ 可以定义为:

$$\beta(n_t)=\begin{cases}0, & n_t<T_s\\ 10, & n_t\in[T_s,T_m]\\ 20, & n_t>T_m\end{cases} \tag{3}$$

显然, 现有研究主要针对的是以金融交易数据为主的公有链, 鲜见面向联盟链和私有链的数据过滤技术研究. 目前, 评价数据过滤技术的标准主要有四点, 包括 1) 过滤质量: 不良信息上链在计算开销和经济成本上都是不可行的, 可有效实现最大限度的数据过滤; 2) 可配置性: 实施数据过滤技术不需要改变运行中的区块链系统, 或者只有较低程度的系统改变; 3) 高可用性: 数据过滤过程不妨碍区块链的正常运行; 4) 低负载性: 数据过滤不会使得区块链系统的性能受到严重降低.

7 数据隐藏技术

区块链数据的公开透明性是其降低系统节点的信任成本、消除信息优势、实现数据共享的重要基础. 然而在特定场景下, 链上数据也存在明显的数据隐藏的需求. 例如, 用户个人信息上链后, 需要在合适的场景下显示合适的信息, 例如在交易购物时需要显示其银行账号信息, 而在浏览信息时则不需要显示; 另外, 不良信息上链后在特定场景下也需要加以隐藏, 使其不可读取, 以避免产生负面影响. 这在目前的区块链上是难以实现的. 现有文献中, 仅有零星讨论涉及到区块链数据的隐藏技术.

第一种方案是以加密方式在链上存储数据的密文, 并向授权用户或者 DAPP 开放解密密钥. 这种方式相对灵活可控, 但缺点是难以在多个实体之间有效地管理密钥, 可能会出现由于密钥丢失或者泄露而导致隐私数据永远不可访问或者完全公开. 此外, 迅猛发展的量子计算也是潜在的威胁.

第二种方案是利用第 3 节所述的链上数据修改技术, 将需要隐藏的数据替换为适合开放的其他数据. 然而, 这种方式通常需要修改区块链协议来集成变色龙哈希、秘密共享等较为复杂的密码学工具, 因而在大规模区块链网络环境下可能存在性能瓶颈.

第三种方案是链下存储实际数据、链上仅存储时间戳和指向链下数据的哈希指针. 当链下数据需要修改或者删除时, 该数据在特定时间点存在的事实将会保留在区块链中, 而且仅利用链上哈希数据无法重构原始数据. 这种方案的优势是较好地解决

了区块链不可篡改性与数据隐藏需求的冲突, 然而其缺陷是可能会降低区块链安全性并引入更多的攻击行为.

值得一提的是, 上述方案大多是思路性的讨论, 尚未有成熟的系统或者应用验证其有效性. 因此, 链上数据隐藏技术仍需进一步的研究和探索.

8　亟需研究的关键问题

本节将总结可编辑区块链领域迫切需要研究解决的若干关键问题.

8.1　可编辑性与安全可信性的兼容

作为一个颇受争议的新热点, 可编辑区块链技术受到研究者质疑最多的就是数据编辑可能为区块链带来中心化安全风险, 并进而降低链上数据的可问责性 (Accountability) 和可信任性 (Trustability), 其中可问责性表示用户可以确知某项链上数据是否经过了修改, 而可信任性则意味着用户确知链上数据经过了所有验证者的共识验证. 显而易见, 现有文献提出的数据修改方案大多存在着中心化的变色龙哈希陷门密钥或者中心化的控制策略 (P) 等缺陷, 可能存在掌握编辑权的恶意用户随意篡改链上数据的安全风险. 因此, 如何实现去中心化的编辑权是兼顾安全性和可编辑性的核心问题.

从技术角度来讲, 这意味着验证者通过原来的 PoW、PoS (Proof-of-stake, 权益证明) 等数据验证共识过程实现记账权的去中心化之外, 还需要增加相对独立的第二层共识过程以实现编辑权的去中心化. 如何设计适用于第二层共识的去中心化算法, 这种双层共识机制如何相互协调交互, 如何保障双层共识的安全性等都是目前悬而未决的关键问题, 迫切需要研究者攻关解决.

8.2　上下文相关的交易级编辑技术

在加密货币型公有链中, 数据的主要表现形式是上下文相关和高度依赖的交易链条, 并在数据结构上呈现出以铸币交易 (如 Coinbase) 为顶点、以加密货币交易为边的树状结构. 如果存在错误交易或者非法交易, 则可能导致其后续交易出现不一致性, 因而必须修改或删除该无效交易并重构以该交易为根节点的交易子树. 例如, 在图 7 中, $A1$ 和 $B3$ 是铸币交易生成的 UTXO, 假设交易 $B2 \to C2$ 是错误或非法交易, 则以 $C2$ 为顶点的后续交易都有可能受到影响. 现有文献中, 除文献 [23] 稍作讨论之外, 其他研究均假设其数据编辑不涉及这种上下文相关的交易, 因此仅需修改交易 $B2 \to C2$ 即可. 实际上, 这种交易链重构是实现加密货币型区块链的数据编辑的关键技术, 迫切需要进一步研究.

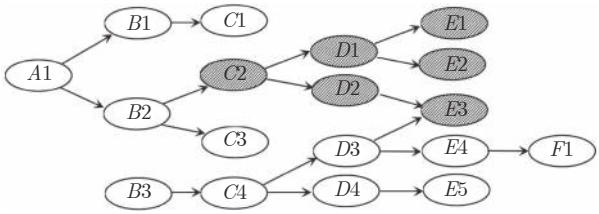

图 7　上下文相关的交易级编辑与重构示例

Fig. 7　An illustrative example of transaction-level editing and reconstruction

形式上, 假设以某错误交易为顶点的交易子树共有 N 个后续交易节点, 涉及到 M 个用户 ($M \leq N$), 则针对此错误交易的数据编辑过程必须同时重构其交易子树, 且重构过程必须保证该子树中所交易的加密货币数量守恒, 即不能凭空新增或者销毁. 以图 7 为例, 错误交易子树共 $N = 6$ 个节点 $\{C2, D1, D2, E1, E2, E3\}$, 由于部分节点可能存在对应同一用户的情况, 因此 $M \leq 6$. 在此基础上, 本文认为一种可行的交易树重构方法为: 1) 首先由 M 个用户之一 (如受损失方) 提出交易修改请求, 定位该错误交易和涉及到的全部 M 个用户、以及子树重构方案; 2) 冻结该错误交易子树上所有的加密货币; 3) M 个用户针对新的子树重构方案做投票共识; 需要说明的是, 这里是区块链用户的共识而非矿工的共识, 因此共识过程可能通过链下谈判达成. 重构方案可以是该子树全部被剪枝, 也可以是子树上所有加密货币在 M 个用户上的重新分配方案; 4) 达成共识后, M 个用户共同签名将子树重构方案发送给验证者; 5) 验证者验证重构方案正确且无一致性冲突后, 采用第 3 节所述方案修改错误交易及其子树, 即可保证交易链条的一致性.

8.3　面向编辑权的冲突与竞争

可编辑区块链的冲突消解机制也是迫切需要研究的课题, 旨在保证链上数据始终一致、稳定、安全和可信. 主要体现在如下两方面:

一方面是数据编辑操作可能带来的内容冲突. 区块链的数据都是经过验证者共识过程后上链的, 而这些数据被修改后生成的新数据一般只是经过逻辑或语法层面的简单验证, 因此数据编辑操作可能会导致链上数据的前后不一致性或者内容冲突. 因此, 必须设计一种链上数据的一致性和冲突检测机制, 使得用户在编辑数据之前即可确知其操作是否会引发链上冲突.

另一方面是竞争编辑权可能带来的用户群体博弈. 区块链系统内的用户和验证者在编辑链上数据之前, 必须首先针对此次编辑操作以及编辑后的新

数据达成有效且稳定的共识. 否则, 当针对链上数据存在不同的意见群体和舆论场时, 将会出现双方或者多方竞争数据编辑权限, 导致链上数据被轮流、频繁地修改, 从而降低链上数据的稳定性、安全性和可信性.

8.4 具有普适性的可编辑区块链技术

由于比特币或以太坊等加密货币型公有链具有更好的不可篡改性, 数据编辑的技术难度更高, 因而现有研究大多是以比特币等公有链为原型和实验环境加以研究的, 其研究方法高度依赖于比特币的数据结构和底层区块链结构, 与联盟链和私有链等其他类型的区块链结构的适配度较低, 与 PoW 之外的共识机制的适配度也较低, 这使得现有研究的可扩展性和可实现性有待提高. 尽管研究者已经针对 PoS 共识和物联网等场景开展了探索性的研究[44−46], 更为普适性的编辑技术仍有迫切的需求.

8.5 链上数据的内容安全与监管

现阶段, 防治不良信息上链是可编辑区块链技术的重要应用场景, 存在迫切的国家信息安全与监管需求. 因此, 链上数据的内容安全与监管是未来的重要研究课题. 本文已经从微观角度详细阐述了区块链数据编辑的技术和方法; 然而, 实际上更为重要的是宏观层面上可信区块链生态体系的建设和治理.

首先是加强区块链信息的内容安全性核查与监管. 针对主流区块链技术和平台, 迫切需要研究区块链的信息源核查技术, 特别是区块链的信源评估方法、上链信息核查方法、基于内容安全的共识算法等, 从源头上保证上链信息的真实性和合理性;

其次, 针对已上链信息, 需要研究区块链大数据的深度分析和安全预警技术, 实现对区块链数据的常态化安全巡查、有害信息的精准定位与深度分析等, 在保持区块链技术极难篡改特点的同时, 提高区块链对有害信息的自动评估与自我净化能力.

最后是加强网络舆论的预测与引导. 区块链本质上仅是信息存储和共享的载体和工具, 解决其内容安全问题的 “功夫在链外”. 因此, 需要研究和发展网络舆论大数据的实时采集和深度分析手段, 以海量社会传感器网络为基础, 以知识自动化技术为核心, 通过对社会态势的全面感知、建模、预测、决策与引导, 将社会公众的问题和矛盾化解在萌芽阶段, 并推动传统社会管理模式向分布式、集约化、智能化、全响应的智慧社会管理模式的转型[47−50].

9 结束语

近年来, 区块链的去中心化和不可篡改等理想特性与现实场景和应用需求之间的对立与冲突呈现出越来越明显的趋势. 这也导致迄今为止区块链技术的 “杀手级应用” 仍然是以比特币和以太坊等局限于纯虚拟经济体系的加密货币, 现实社会中解决实际需求的区块链系统虽为数众多、百花齐放[51−53], 但均是小规模探索和尝试. 究其原因, 去中心化和不可篡改等理想特性在为区块链奠定坚实的数据安全与信任基础的同时, 也在一定程度上限制了区块链脱虚向实、实现大规模产业落地应用的可能性. 因此, 区块链技术正在从理想回归现实. 这种回归是区块链技术的进化还是退化, 将会是未来一段时间内研究者讨论的热点.

技术发展的核心驱动力和根本目标是满足应用需求、解决实际问题. 从这个角度来说, 可编辑区块链技术由国家监管和技术发展双重需求驱动, 虽然现阶段仅有小规模的零星探索性研究, 但有着丰富的研究问题和技术挑战, 吸引着越来越多的研究者从事此领域的科研攻关. 本文给出了可编辑区块链的研究框架, 详细阐述了基于可编辑区块链的数据修改、删除、插入、过滤和隐藏技术, 梳理了未来亟需研究的热点问题, 以期为未来的研究提供有益的启发与借鉴.

致谢

感谢本文写作过程中与西安电子科技大学裴庆祺教授、刘雪峰教授和华东师范大学金澈清教授的有益讨论, 这些讨论为此文提供了许多思路和启发.

References

1 Nakamoto S. Bitcoin: a peer-to-peer electronic cash system [Online], available: https://bitcoin.org/bitcoin.pdf, January 1, 2009

2 Yuan Yong, Wang Fei-Yue. Blockchain: the state of the art and future trends. *Acta Automatica Sinica*, 2016, **42**(4): 481−494
(袁勇, 王飞跃. 区块链技术发展现状与展望. 自动化学报, 2016, **42**(4): 481−494)

3 Yuan Yong, Wang Fei-Yue. *Blockchain Theory and Method*. Beijing: Tsinghua University Press, 2019
(袁勇, 王飞跃. 区块链理论与方法. 北京: 清华大学出版社, 2019)

4 Yuan Yong, Zhou Tao, Zhou Ao-Ying, Duan Yong-Chao, Wang Fei-Yue. Blockchain technology: from data intelligence to knowledge automation. *Acta Automatica Sinica*, 2017, **43**(9): 1485−1490
(袁勇, 周涛, 周傲英, 段永朝, 王飞跃. 区块链技术: 从数据智能到知识自动化. 自动化学报, 2017, **43**(9): 1485−1490)

5 Yuan Y, Wang F Y. Blockchain and cryptocurrencies: model, techniques, and applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, **48**(9): 1421−1428

6 Yuan Y, Wang F Y, Rong C M, Stavrou A, Zhang J, Tang Q, et al. Guest editorial special issue on blockchain and economic knowledge automation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020, **50**(6): 2−8

7 Han Xuan, Yuan Yong, Wang Fei-Yue. Security problems on blockchain: the state of the art and future trends. *Acta Auto-*

matica Sinica, 2019, **45**(1): 206−225
(韩璇, 袁勇, 王飞跃. 区块链安全问题: 研究现状与展望. 自动化学报, 2019, **45**(1): 206−225)

8 Yuan Yong, Ni Xiao-Chun, Zeng Shuai, Wang Fei-Yue. Blockchain consensus algorithms: the state of the art and future trends. *Acta Automatica Sinica*, 2018, **44**(11): 2011−2022
(袁勇, 倪晓春, 曾帅, 王飞跃. 区块链共识算法的发展现状与展望. 自动化学报, 2018, **44**(11): 2011−2022)

9 Truong N B, Sun K, Lee G M, Guo Y K. GDPR-compliant personal data management: a blockchain-based solution [Online], available: https://arxiv.org/pdf/1904.03038.pdf, January 1, 2019

10 Ouyang Li-Wei, Wang Shuai, Yuan Yong, Ni Xiao-Chun, Wang Fei-Yue. Smart contracts: architecture and research progresses. *Acta Automatica Sinica*, 2019, **45**(3): 445−457
(欧阳丽炜, 王帅, 袁勇, 倪晓春, 王飞跃. 智能合约: 架构及进展. 自动化学报, 2019, **45**(3): 445−457)

11 Wang S, Ouyang L W, Yuan Y, Ni X C, Han X, Wang F Y. Blockchain-enabled smart contracts: architecture, applications, and future trends. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019, **49**(11): 2266−2277

12 Zeng Shuai, Yuan Yong, Ni Xiao-Chun, Wang Fei-Yue. Scaling blockchain towards Bitcoin: key technologies, constraints and related issues. *Acta Automatica Sinica*, 2019, **45**(6): 1015−1030
(曾帅, 袁勇, 倪晓春, 王飞跃. 面向比特币的区块链扩容: 关键技术, 制约因素与衍生问题. 自动化学报, 2019, **45**(6): 1015−1030)

13 Garay J, Kiayias A, Leonardos N. The Bitcoin backbone protocol: analysis and applications. In: Proceedings of the 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques. Sofia, Bulgaria: Springer, 2015. 281-310

14 Ateniese G, Magri B, Venturi D, Andrade E. Redactable blockchain - or - rewriting history in Bitcoin and friends. In: Proceedings of the 2017 IEEE European Symposium on Security and Privacy. Paris, France: IEEE, 2017. 111−126

15 Krawczyk H M, Rabin T D. Chameleon hashing and signatures. U.S. Patent 6108783, August 2000

16 Krawczyk H, Rabin T. Chameleon signatures. In: Proceedings of Network and Distributed System Security Symposium. San Diego, CA, USA: Internet Society, 2000. 143−154

17 Li Pei-Li, Xu Hai-Xia, Ma Tian-Jun, Mu Yong-Heng. Research on fault-correcting blockchain technology. *Journal of Cryptologic Research*, 2018, **5**(5): 501−509
(李佩丽, 徐海霞, 马添军, 穆永恒. 可更改区块链技术研究. 密码学报, 2018, **5**(5): 501−509)

18 Rajasekhar K, Yalavarthy S H, Mullapudi S, Gowtham M. Redactable blockchain and it's implementation in bitcoin. *International Journal of Engineering & Technology*, 2018, **7**(1.1): 401−405

19 Ashritha K, Sindhu M, Lakshmy K V. Redactable blockchain using enhanced chameleon hash function. In: Proceedings of the 5th International Conference on Advanced Computing & Communication Systems (ICACCS). Coimbatore, India: IEEE, 2019

20 Shamir A. How to share a secret. *Communications of the ACM*, 1979, **24**(11): 612−613

21 Camenisch J, Derler D, Krenn S, Pöhls H C, Samelin K, Slamanig D. Chameleon-hashes with ephemeral trapdoors. In: Proceedings of the 20th IACR International Conference on Practice and Theory in Public-Key Cryptography (PKC). Amsterdam, the Netherlands: Springer, 2017. 152−182

22 Derler D, Samelin K, Slamanig D, Striecks C. Fine-grained and controlled rewriting in blockchains: chameleon-hashing gone attributable-based. In: Proceedings of the 26th Network and Distributed Systems Security (NDSS). San Diego, USA, 2019

23 Puddu I, Dmitrienko A, Capkun S. *u*chain: how to forget without hard forks. Cryptology ePrint archive: report 2017/106 [Online], Available: http://eprint.iacr.org/2017/106, January 1, 2019

24 Politou E, Casino F, Alepis E, Patsakis C. Blockchain mutability: challenges and proposed solutions. *IEEE Transactions on Emerging Topics in Computing*, 2019

25 Marsalek A, Zefferer T. A correctable public blockchain. In: Proceedings of the 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE). Rotorua, New Zealand: IEEE, 2019

26 Deuber D, Magri B, Thyagarajan S A K. Redactable blockchain in the permissionless setting. In: Proceedings of the 2019 IEEE Symposium on Security and Privacy. San Francisco, USA: IEEE, 2019

27 Yuan Yong, Wang Fei-Yue. Parallel clockchain: concept, methods and issues. *Acta Automatica Sinica*, 2017, **43**(10): 1703−1712
(袁勇, 王飞跃. 平行区块链: 概念、方法与内涵解析. 自动化学报, 2017, **43**(10): 1703−1712)

28 Wang F Y, Yuan Y, Rong C M, Zhang J J. Parallel blockchain: an architecture for CPSS-based smart societies. *IEEE Transactions on Computational Social Systems*, 2018, **5**(2): 303−310

29 Qin R, Yuan Y, Wang F Y. Research on the selection strategies of blockchain mining pools. *IEEE Transactions on Computational Social Systems*, 2018, **5**(3): 748−757

30 Qin R, Yuan Y, Wang F Y. A novel hybrid share reporting strategy for blockchain miners in PPLNS pools. *Decision Support Systems*, 2019, (118): 91−101

31 Qin R, Yuan Y, Wang S, Wang F Y. Economic issues in Bitcoin mining and blockchain research. In: Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV). Changshu, China: IEEE, 2018. 268−273

32 Palm E, Schelén O, Bodin U. Selective blockchain transaction pruning and state derivability. In: Proceedings of the 2018 Crypto Valley Conference on Blockchain Technology (CVCBT). Zug, Switzerland: IEEE, 2018

33 Chepurnoy A, Larangeira M, Ojiganov A. Rollerchain, a blockchain with safely pruneable full blocks [Online], Available: https://arxiv.org/pdf/1603.07926, January 1, 2019

34 Florian M, Beaucamp S, Henningsen S, Scheuermann B. Erasing data from blockchain nodes [Online], Available: https://arxiv.org/pdf/1904.08901.pdf, January 1, 2019

35 Bruce J D. The mini-blockchain scheme [Online], Available: http://cryptonite.info/files/mbc-scheme-rev3.pdf, January 1, 2019

36 Feng X Q, Ma J F, Miao Y B, Meng Q, Liu X M, et al. Pruneable sharding-based blockchain protocol. *Peer-to-Peer Networking and Applications*, 2018, **12**(4): 934−950

37 Ethereum White Paper. A next-generation smart contract and decentralized application platform [Online], available: https://github.com/ethereum/wiki/wiki/White-Paper, November 12, 2015

38 Matzutt R, Hiller J, Henze M, Ziegeldorf J H, Müllmann D, Hohlfeld O, et al. A quantitative analysis of the impact of arbitrary blockchain content on Bitcoin. In: Proceedings of the 22nd International Conference on Financial Cryptography and Data Security (FC). Springer, 2018

39 Sward A, Vecna I, Stonedahl F. Data insertion in Bitcoin's blockchain. *Ledger*, 2018: 3

40 Bartoletti M, Pompianu L. An analysis of Bitcoin OP_RETURN metadata. In: Proceedings of the 4th Workshop on Bitcoin and Blockchain Research. Malta, 2017

41 Wang S, Ding W W, Li J J, Yuan Y, Ouyang L W, Wang F Y.

Decentralized autonomous organizations: concept, model, and applications. *IEEE Transactions on Computational Social Systems*, 2019, **6**(5): 870−878

42 Li J J, Yuan Y, Wang F Y. A novel GSP auction mechanism for ranking Bitcoin transactions in blockchain mining. *Decision Support Systems*, 2019, **124**: 113094

43 Matzutt R, Henze M, Ziegeldorf J H, Hiller J, Wehrle K. Thwarting unwanted blockchain content insertion. In: Proceedings of the 2018 IEEE International Conference on Cloud Engineering (IC2E). Orlando, USA: IEEE, 2018. 364−370

44 Xu J, Li X Y, Yin L Y, Guo B Y, Feng H, Zhang Z F. Redactable proof-of-stake blockchain with fast confirmation, IACR Cryptology ePrint Archive, 2019, **2019**: 1110

45 Huang K, Zhang X S, Mu Y, Wang X F, Yang G M, Du X J, et al. Building Redactable consortium blockchain for industrial Internet-of-Things. *IEEE Transactions on Industrial Informatics*, 2019, **15**(6): 3670−3679

46 Huang K, Zhang X S, Mu Y, Rezaeibagha F, Du X J, Guizani N. Achieving intelligent trust-layer for internet-of-things via self-redactable blockchain. *IEEE Transactions on Industrial Informatics*, 2020, **16**(4): 2677−2686

47 Wang Fei-Yue, Wang Xiao, Yuan Yong, Wang Tao, Lin Yi-Lun. Social computing and computational societies: the foundation and consequence of smart societies. *Chinese Science Bulletin*, 2015, **60**(5−6): 460−469
(王飞跃, 王晓, 袁勇, 王涛, 林懿伦. 社会计算与计算社会: 智慧社会的基础与必然. 科学通报, 2015, **60**(5−6): 460−469)

48 Wang X, Li L X, Yuan Y, Ye P J, Wang F Y. ACP-based social computing and parallel intelligence: societies 5.0 and beyond. *CAAI Transactions on Intelligence Technology*, 2016, **4**(1): 377−393

49 Wang F Y, Yuan Y, Zhang J, Qin R, Smith M H. Blockchainized Internet of Minds: A new opportunity for cyber-physical-social systems. *IEEE Transactions on Computational Social Systems*, 2018, **5**(4): 897−906

50 Ding W W, Wang S, Li J J, Yuan Y, Ouyang L W, Wang F Y. Decentralized autonomous organizations:the state of the art,analysis framework and future trends. *Chinese Journal of Intelligent Science and Technologies*, 2019, **1**(2): 202−213
(丁文文, 王帅, 李娟娟, 袁勇, 欧阳丽炜, 王飞跃. 去中心化自治组织: 发展现状、分析框架与未来趋势. 智能科学与技术学报, 2019, **1**(2): 202−213)

51 Zhang Jun, Yuan Yong, Wang Xiao, Wang Fei-Yue. Quantum blockchain: Can blockchain integrated with quantum information technology resist quantum supremacy? *Chinese Journal of Intelligent Science and Technologies*, 2019, **1**(4): 409−414
(张俊, 袁勇, 王晓, 王飞跃. 量子区块链: 融合量子信息技术的区块链能否抵御量子霸权? 智能科学与技术学报, 2019, **1**(4): 409−414)

52 Ouyang Li-Wei, Yuan Yong, Zhang Jun, Wang Fei-Yue. A novel blockchain-based surveillance and early-warning technology for infectious diseases. *Chinese Journal of Intelligent Science and Technology*, 2020, **2**(2): 129−137
(欧阳丽炜, 袁勇, 张俊, 王飞跃. 基于区块链的传染病监测与预警技术. 智能科学与技术学报, 2020, **2**(2): 129−137)

53 Wang S, Wang J, Wang X, Qiu T Y, Yuan Y, Ouyang L W, et al. Blockchain powered parallel healthcare systems based on the ACP approach. *IEEE Transactions on Computational Social Systems*, 2018, **5**(4): 942−950

袁 勇 中国科学院自动化研究所复杂系统管理与控制国家重点实验室副研究员, 中国自动化学会区块链专委会主任. 2008 年获得山东科技大学计算机软件与理论专业博士学位. 主要研究方向为社会计算, 计算广告学与区块链. 本文通信作者.
E-mail: yong.yuan@ia.ac.cn
(**YUAN Yong** Associate professor at the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences. He is also the director of Technical Committee on Blockchain, Chinese Association of Automation. He received his Ph. D. degree in computer software and theory from Shandong University of Science and Technology in 2008. His research interest covers social computing, computational advertising, and blockchain. Corresponding author of this paper.)

王飞跃 中国科学院自动化研究所复杂系统管理与控制国家重点实验室主任, 国防科技大学军事计算实验与平行系统技术研究中心主任, 中国科学院大学中国经济与社会安全研究中心主任, 青岛智能产业技术研究院院长. 主要研究方向为平行系统的方法与应用, 社会计算, 平行智能以及知识自动化.
E-mail: feiyue.wang@ia.ac.cn
(**WANG Fei-Yue** Director of the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences. Professor of the Research Center for Computational Experiments and Parallel Systems Technology, National University of Defense Technology. Director of China Economic and Social Security Research Center in University of Chinese Academy of Sciences. Dean of Qingdao Academy of Intelligent Industries. His research interest covers methods and applications for parallel systems, social computing, parallel intelligence, and knowledge automation.)