计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2021.20210804
58(10): 2099-2119, 2021

区块链数据隐私保护:研究现状与展望

王晨旭[1,3] 程加成[1] 桑新欣[1] 李国栋[2] 管晓宏[3]
[1](西安交通大学软件学院 西安 710049)
[2](西安交通大学网络信息中心 西安 710049)
[3](智能网络与网络安全教育部重点实验室(西安交通大学) 西安 710049)
(cxwang@mail.xjtu.edu.cn)

Data Privacy-Preserving for Blockchain: State of the Art and Trends

Wang Chenxu[1,3], Cheng Jiacheng[1], Sang Xinxin[1], Li Guodong[2], and Guan Xiaohong[3]
[1](*School of Software Engineering, Xi'an Jiao Tong University, Xi'an* 710049)
[2](*Network Information Center, Xi'an Jiao Tong University, Xi'an* 710049)
[3](*Key Laboratory of Intelligent Network and Network Security (Xi'an Jiao Tong University), Ministry of Education, Xi'an* 710049)

Abstract As a distributed ledger, blockchain solves the decentralized trust problem by integrating a series of techniques such as distributed consensus, P2P (Peer to Peer) network, smart contracts, and cryptography. Blockchain has a meaningful impact on the society and lifts a bloom of researches and applications due to the characteristics of immutability and decentralization. Blockchain technology has a broad scope of applications, and its unique advantages can deal with the pain points in many industry scenarios. However, the blockchain technology is faced with the problem of data privacy leakage in its applications, such as the disclosure of transaction, account and personal information privacy, which greatly impose restrictions on the application scope and fields. Data privacy-preserving for blockchain has become one of the key problems concerned by researchers. In this survey, we first describe the evolutionary history of blockchain technology, define the concept of corresponding privacy according to applications in the field of blockchain and introduce the main technical points and the technology architecture of blockchain. Then we summarize the privacy-preserving problems faced by the blockchain technology and explore the existing solutions based on the proposed concept of data privacy protection. Finally, some problems that still need to be addressed and future research directions of data privacy-preserving for blockchain are discussed based on the analysis.

Key words blockchain; privacy-preserving; side-channel attacks; coin mixing; channel isolation

摘 要 区块链作为一种分布式账本,集成了分布式共识、对等(peer to peer, P2P)网络、智能合约及密码学等技术,解决了"去中心化"的信任问题.区块链凭借其不可篡改、去中心化等特性,对社会各个领域产生了深远影响,掀起了区块链技术的研究与应用热潮.区块链技术应用场景十分广泛,其独特优势

收稿日期:2021-08-13;**修回日期**:2021-08-19
基金项目:国家自然科学基金项目(61602370);陕西省自然科学基金项目(2021JM-018);深圳基础研究基金项目(JCYJ20170816100819428);中央高校基本科研业务费专项资金(1191320006)
This work was supported by the National Natural Science Foundation of China (61602370), the Natural Science Foundation of Shanxi Province (2021JM-018), Shenzhen Fundamental Research Program (JCYJ20170816100819428), and the Fundamental Research Funds for the Central Universities (1191320006).

能够解决许多行业中的痛点.但是,区块链技术在应用过程中面临着数据隐私泄露的问题,极大地限制了区块链的应用范围和领域,区块链数据隐私保护方案已成为研究者关注的重点问题之一.基于数据隐私保护的基本概念,详细分析了区块链各技术要点面临的隐私泄露问题,探索并总结了当前区块链数据隐私保护的解决方案.最后,结合目前区块链数据隐私保护研究的最新进展,对未来区块链数据隐私保护的研究方向进行了展望.

关键词 区块链;隐私保护;旁道攻击;混币交易;通道隔离

中图法分类号 TP391

区块链作为一种分布式账本技术,实现了"去中心化"的信任,其核心思想及技术体系最早见于中本聪 2008 年发表的比特币白皮书[1].随着区块链技术不断演进,逐步集成了分布式数据库[2]、共识机制[3]、P2P 网络[4]、智能合约[5]及密码学[6]等技术.区块链技术凭借其不可篡改、分布式容错、去中心化等特性,迅速得到学术界和工业界的关注,被认为是继移动互联网后的第 5 代互联网颠覆性技术[7],对社会各个领域产生了深远影响.如今,数据已成为宝贵的资源,企业可以基于收集的数据预测未来趋势、优化决策过程、为用户提供个性化服务等[8],但同时带来了数据隐私泄露的问题.随着区块链技术应用领域的不断扩展,其面临的数据隐私泄露问题逐渐成为大家的关注重点.

2009 年比特币创世区块的诞生标志着催生区块链技术的比特币项目正式落地,随后区块链在金融领域以数字货币的形式迅速发展,衍生出类似比特币的莱特币[9]、夸克币[10]、无限币[11]等币种,主要应用于货币支付.区块链技术在数字货币领域的应用阶段称为区块链 1.0 阶段,此时区块链架构主要围绕数字货币的发行与交易支付等需求而设计,在区块链技术的实际应用中,由于需要保证账本内容的一致性、可溯源、可验证性,账本内容需要对区块链网络中的所有节点公开.因此,比特币账本透明性导致的数据隐私泄露问题最初并未得到广泛关注.区块链技术为保证系统安全性及可验证性需要公开账本内容,导致恶意节点能够获取所有账本信息,通过聚类技术分析账本信息可以窥探区块链匿名账户与现实用户之间的身份关联关系.目前,数据隐私泄露已成为比特币及其延伸项目面临的重要潜在问题.

智能合约概念的引入极大增强了区块链技术的灵活性,拓展了其应用的范围,标志着区块链技术进入了 2.0 阶段.智能合约与数字货币的紧密结合在金融领域产生了更加广泛的应用前景,同时带来了一定程度的数据隐私泄露问题,例如恶意攻击者可以通过网络中安全性较为薄弱的节点监听智能合约的执行情况,窃取用户隐私信息等.通道技术及零知识证明技术[12]为实现区块链数据隐私保护提供了可行的解决方案.2014 年 Zerocoin 协议[13]演进为 Zerocash 系统[14],并于 2016 年构建成 Zcash 加密货币系统[15].一系列侧重隐私保护的加密货币的不断改进,侧面反映了区块链社区对数据隐私泄露问题的重视程度不断提高.简洁非交互式零知识证明(zero-knowledge succinct non-interactive argument of knowledge, ZK-SNARK)[16]、环签名[17]等密码学技术实现了一定程度上的数据隐私保护,但仍存在数据隐私泄露的风险,如采用远程旁道攻击可探测 Zcash 等加密货币系统的交易金额、用户 IP 地址等隐私信息[18].智能合约的应用及区块链相关技术导致的数据隐私问题已经引起业内研究与开发人员的广泛关注,数据隐私泄露问题成为区块链系统应用开发面临的主要问题之一.

区块链技术是具有普适性的底层技术框架,可以为金融、经济、科技甚至政务等各领域带来深刻变革[19].区块链在金融行业之外其他领域的应用标志着区块链发展进入 3.0 阶段,该阶段需要区块链平台能够满足更加复杂的商业逻辑及企业间合作时的数据隐私保护需求.数据隐私泄露极有可能给企业和个人带来致命性打击.因此,在区块链系统实际落地前,提出合适的方案解决数据隐私泄露问题是必不可少的一环.为满足更高程度数据隐私保护需求,Hyperledger Fabric[20]采用通道隔离机制提供一定程度的数据隐私保护功能,FISCO BCOS[21]则通过配置环签名、群签名、同态加密等密码学工具实现数据隐私保护.目前针对链上数据及智能合约的攻击方法也在逐步增加,区块链数据隐私泄露的问题已经成为区块链架构设计人员及开发人员的重要考虑因素.

区块链在不同发展阶段的成果相互影响促进了区块链技术的发展,同时区块链技术与社会各个领域的结合越发紧密,诸多传统数据隐私保护方案在

区块链应用中的不适用性和源于区块链技术自身体系结构所导致的数据隐私泄露问题已经逐步暴露出来.为此,本文系统整理了区块链历史发展中遇到的数据隐私泄露问题及解决方案,提出针对区块链系统应用需求的隐私定义及分类,期望为当前及未来区块链技术数据隐私保护方向的发展研究提供参考.

1 区块链及数据隐私保护概述

区块链技术最初应用于互联网金融领域,旨在解决第三方中心机构的信任问题,为去中心化的交易提供一种信任机制.区块链技术的核心思想源于2009年发布实施的比特币项目,该项目涉及的核心技术主要有密码学、激励机制、工作量证明机制、P2P网络、分布式数据库[1]等技术.该项目简化了交易流程,达到降低交易成本的目的,实现去中心化的交易体系.

近年来,区块链数据隐私泄露问题已经引起相关研究者的广泛关注,关于技术和隐私的关系理论可以追溯到19世纪90年代[22].根据中国区块链技术产业和发展论坛对区块链隐私的标准定义,隐私指仅与个人利益相关且不需要强制公开的个人信息及个人领域.隐私的主体是自然人,客体是个人信息和个人领域,内容指特定个人对信息和领域的秘而不宣,不愿第三方探知和干涉的事实和行为[23].根据隐私定义,用户的隐私数据内容可分为2类:

1) 个人信息.直接与用户真实身份信息相关,例如用户的真实姓名、电话、年龄、实际住址等隐私信息.

2) 个人领域.间接与用户真实身份信息相关,例如用户的公私钥对、IP地址、交易内容、节点位置及网络中用户收款地址与现实中用户身份的关联关系等隐私信息.

保护以上2类用户信息的机密性是区块链数据隐私保护技术的首要目标,区块链的技术要点中存在一定程度的数据隐私泄露问题,目前针对各技术的数据隐私泄露问题研究者陆续提出了相应的解决方案.

分布式数据库技术是区块链核心技术之一,是一种实现数据库在逻辑上统一但物理上分散的技术[24],具有数据透明性、一致性、冗余性等特性.对于比特币及其延伸项目,分布式数据库中存储的交易内容公开透明,可由整个系统中的用户访问,该特性能够保证此类区块链系统具有良好监管性.但这一技术也导致了比特币系统的匿名性不能真正意义上实现用户身份的匿名化,攻击者利用账本的公开透明性可以侵犯用户的账户隐私及交易隐私,如针对账户交易模型可通过账户聚类技术统计出网络中的账号地址和现实生活中用户的身份信息之间的关联,针对UTXO交易模型可通过交易聚类、交易溯源等技术统计出交易之间的关联性以及交易和地址之间的关联性.

P2P技术为去中心化系统中各对等节点对账本内容达成共识提供了桥梁,由对等节点组成的P2P网络有其独特优势,网络中的节点功能及地位对等,可以同时处于服务者与被服务者2种角色状态.集成了P2P技术的区块链系统实现了去中心化的信任,具有很强的健壮性和可扩展性,但由于其去中心化的架构体系,P2P网络中存在节点自身信息隐私泄露及节点之间通信信息隐私泄露两大问题.P2P网络中用户节点的性能及安全性通常很难保持一致,因此攻击者可以通过攻击网络中安全性较为薄弱的节点获取网络中的敏感信息从而窥探用户隐私.攻击者可以通过在P2P网络中设置恶意监听节点、监听及分析节点之间的通信信息,进一步分析出节点之间的通信数据内容,从而损害整个区块链网络的隐私性.

以太坊智能合约的出现对区块链应用场景的扩展具有划时代意义,其应用范围从数字货币领域拓展到了通用计算领域,开发者已经可以使用编程语言设计一些复杂的业务逻辑满足场景需求.在某些应用场景下智能合约本身需要一定的隐私性,例如借贷应用场景中智能合约需要确保自身的保密性,其对数据的操作有可能被攻击者观察分析并利用统计学方法推测出交易内容[25],因此智能合约本身的泄露也可能导致攻击者掌握用户交易的隐私情况.密码学的安全技术与智能合约结合构成了隐私智能合约的概念,例如将密码学中的零知识证明、同态加密[26]等技术结合智能合约可以有效实现隐私保护,隐私智能合约能够极大增强用户交易数据的匿名性.此外,智能合约是一种特殊的计算机程序,可能自身存在漏洞,这些漏洞也可能导致隐私数据的泄露.

密码学安全技术为人类社会进入信息化时代奠定了基础,在整个信息技术领域有着举足轻重的地位.区块链体系中融合了大量密码学安全技术的研究成果,例如比特币系统中区块的哈希链式数据结构提供了防篡改的特性,公私钥对密码体系为比特币系统提供了用户匿名性,公钥基础设施(public

key infrastructure, PKI)系为用户身份认证提供了保障,数字签名技术保证了交易内容的不可篡改性和不可抵赖性.比特币借助于密码学相关技术利用"去信任化"的交易模型代替了传统的基于信任的交易模型,实现了不需要第三方中介参与的电子交易系统.密码学安全技术是区块链系统实现数据隐私保护的关键,但密码学技术存在被破解的可能,如果区块链系统所使用的密码学技术被破解,建立在被破解密码学技术上的区块链系统的安全性和隐私性将不复存在,例如 Shor 算法[27]及 Grover 算法[28]对区块链体系的威胁性.此外,未来商用量子计算机对基于密集计算型算法的区块链系统也将带来一定的冲击.目前,基于诱骗态 BB84 算法[29]的量子秘钥分发技术[30]已经走向商用,未来更安全的基于量子纠缠进行量子秘钥分发的技术也已出现,区块链技术应逐步融入针对抗量子计算型的密码学安全技术以保证区块链系统的安全性.

区块链层次结构自上而下依次为:应用层、合约层、激励层、网络层、数据层,其中数据层、网络层和共识层为区块链技术的核心元素.数据层、网络层及应用层与区块链数据隐私威胁问题紧密相连,祝烈煌等人[31]从数据层、网络层及应用层角度出发对数据隐私威胁进行分析并总结出相应的数据隐私保护解决方案.本文针对区块链技术要点及应用场景,对区块链技术的隐私内容进行了细致的分类,依次为链上数据隐私、针对智能合约的链码隐私、针对 P2P 网络的网络隐私、针对跨链技术的跨链隐私、针对实际应用场景的区块链应用隐私,基于以上分类提出了相应的隐私威胁及挑战并给出相应解决方案,表 1 总结了区块链技术面临的隐私问题分类及相关解决方案.最后,结合区块链技术的发展趋势,本文讨论了区块链数据隐私保护未来可能的研究方向.

Table 1 Categories of Blockchain Privacy Issues and Related Technologies for Privacy Preserving

表 1 区块链隐私分类及实现隐私保护的相关技术

区块链隐私类别	子类	隐私内容	威胁、攻击、挑战	解决方案
链上数据隐私	交易隐私	交易发起方、接收方、交易金额等隐私信息	借助爬虫技术爬取账本信息、论坛及交易所等区块链服务信息,构建交易网络拓扑、用户网络拓扑,利用溯源技术及交易特征通过数据分析获取隐私信息	混淆机制、零知识证明、同态加密、环签名、通道隔离、权限限制、承诺方案、基于属性加密、差分隐私
	账户地址隐私	账户地址余额、账户之间交易联系等隐私信息		
	用户身份信息	用户真实姓名、年龄、住址、身份证号等隐私信息		
链码隐私		链码源代码、版本号等隐私信息	人为窃取、控制网络节点、利用链码漏洞获取隐私信息	通道隔离、权限限制、可信硬件执行环境
网络隐私	节点隐私	服务器地理位置、节点物理信息、系统版本、节点 IP 等隐私信息	爬取区块链网络节点公开信息或在区块链网络中部署节点监听	可信第三方转发、混合网络、洋葱路由、大蒜路由、雷电网络、闪电网络
	通信隐私	节点间通信数据明文及密文、通信流量等隐私信息	在区块链网络中部署监听节点进行远程旁道攻击	
跨链隐私		跨链操作时涉及的跨链数据及相关账户地址等隐私信息	在区块链网络中部署监听节点、截取跨链桥的数据信息	隐私智能合约、可信硬件执行环境、去中心化身份识别器
应用隐私	用户端隐私	支付流敏感信息、浏览器 Cookie、密钥存放位置等隐私信息	软件权限申请,收集论坛、服务商等公开信息,秘钥盗取,插件漏洞	发布官方插件、身份认证、普及隐私泄露危险、推广隐私保护方法
	服务端隐私			

2 区块链数据隐私保护问题及挑战

2.1 链上数据隐私及威胁

链上数据隐私包括区块链网络中任何与用户个人信息及个人领域相关的数据信息,共分为 3 类:

1) 交易隐私.交易发起方、接收方、交易金额、用户交易特征等隐私信息.

2) 账户地址隐私.账户地址余额、账户之间交易联系等隐私信息.

3) 用户身份信息.用户真实姓名、年龄、住址、身份证号等隐私信息.

比特币采用基于未花费交易输出集(unspent transaction outputs, UTXO)的交易模型,交易的公开透明性及开放性允许任何加入区块链网络的用户能够轻易获取详细交易内容,同时区块链的链式数据结构以及 Merkle 树结构可以保证系统中发生的每一笔交易都可以轻松溯源.尽管比特币凭借假名机制保证了一定的匿名性,但它仍然存在许多隐私问题.

Reid 等人[32]于 2013 年下载了 2009-01—2011-07 比特币系统的公开账本数据,局部交易网络如图 1

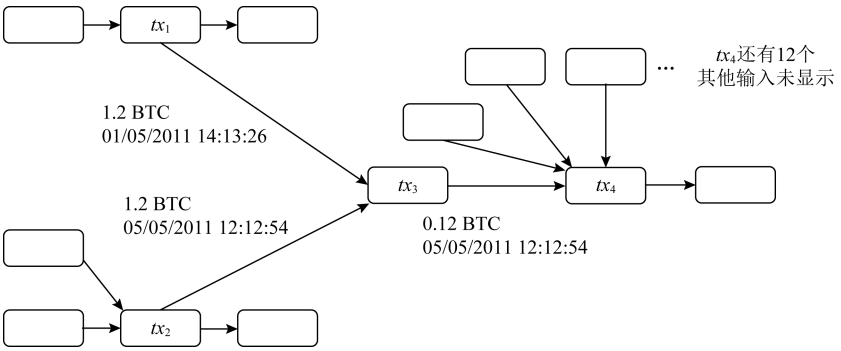

Fig. 1 An example sub-network from the transaction network[32]

图 1 交易网络中子网络示例[32]

所示,网络中的节点表示一次交易,有向边表示交易之间的输出-输入对,每条边同时标记了交易金额以及时间戳,从而构建了比特币的交易网络.假设一个交易的多个输入地址属于同一个用户,基于该假设,可以聚合同一用户的所有地址并构建如图 2 所示的用户网络,菱形表示公钥(地址),圆圈表示不同用户,从而揭示出用户的比特币资金余额及流向.

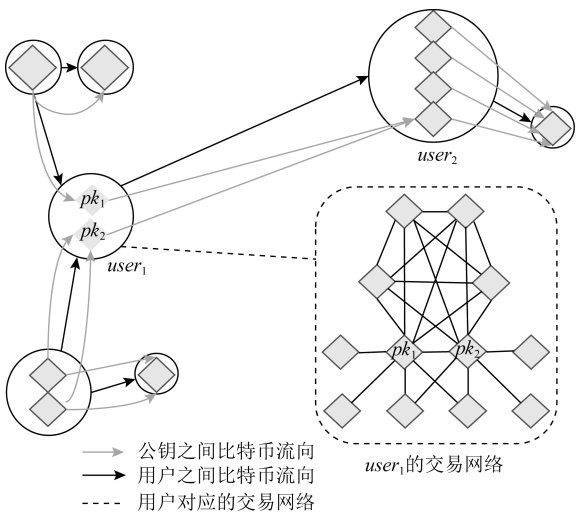

Fig. 2 An example sub-network from the user network[32]

图 2 用户网络局部示例[32]

Androulaki 等人[33]基于真实的比特币系统和一个模拟器模拟大学环境中比特币使用情况,对比特币系统的匿名性进行了评估.该实验有 2 个假设,假设 1:一个交易的多个输入地址属于同一用户;假设 2:找零地址与输入地址属于同一个用户.基于假设 1 与假设 2 对模拟器输出的交易信息进行预处理,将结果提交至聚类算法,实验结果表明近 40% 的用户数据信息可以被揭露出来.Ron 等人[34]下载了比特币的完整历史,通过研究用户行为、比特币输入输出情况,得出其关联交易图的许多统计特性,结果表明用户采取资产转移以保护自身资产隐私的方法不能有效保护个人隐私.此外,论文还分析了系统中所有的大型交易,并发现了一个被人使用长链和分叉操作来精心隐藏的事实——几乎所有交易都与 2010 年 11 月发生的一个大型交易密切相关.

此类假设及攻击方法基于比特币系统自身设计缺陷,通过分析账本内容对比特币系统去匿名化从而获得地址之间关联,从而削弱比特币系统的匿名性.Meiklejohn 等人[35]进一步研究了比特币系统的匿名性,通过使用启发式聚类分析算法对比特币钱包进行分组,然后对用户进行分类,从而识别出同一个用户的不同地址.作者对一些公开地址以及违法犯罪相关案件进行了分析,发现了许多与案件涉及与交易地址相关联的其他地址.2015 年,Monaco 等人[36]基于对比特币长期交易行为的观察,基于行为生物统计学思想和动态网络特征提出了一种识别和验证比特币用户的新方法,实验结果表明随着时间推移用户的匿名性无法得到保证.

2.2 智能合约隐私及威胁

智能合约概念的提出为区块链提供了更广阔的应用场景,但智能合约技术在实际应用时有可能导致数据隐私的泄露.用户发起函数调用后系统会构建智能合约交易,区块链系统中的许多节点会对该交易进行处理,这要求与交易相关的操作和数据需要对所有节点开放,在处理敏感数据的应用场景时会存在隐私泄露的问题,如投票方案、医疗数据收集等[37].如何处理敏感数据与智能合约的关系需要进一步研究.以拍卖合同为例,实现一个基于以太坊的拍卖合同系统并不困难,但是建立一个密封投标机制的拍卖系统并不容易.密封投标拍卖的工作原理大致为:在投标阶段,投标人向拍卖人提交密封投标,

每个投标人的投标金额对其他投标人不可见,随后投标被解封,通过比较投标金额选择获胜者.但是以太坊的所有交易金额都是公开的,如何结合智能合约实现数据隐私保护是目前面临的一个挑战[38].

如果智能合约代码编写不当,则其在执行过程中会引发漏洞[39],进而导致数据隐私的泄露.智能合约作为计算机系统程序的一种特殊形式,本身难以摆脱漏洞[40].智能合约在执行前需要编译成二进制字节码,利用智能合约分析工具,如 Oyente[41] 等开源工具,可以对智能合约的漏洞进行分析.虽然某些系统在设计保护隐私的加密货币方面取得了一定的成果,如 Zerocash[14],Zerocoin[13] 和其他加密货币系统,但这些系统放弃了可编程性,并且尚不清楚如何在不向矿工公开明文中交易和数据的前提下实现可编程性.

智能合约在执行前需要被部署到区块链节点上,由于区块链网络节点的物理分散性,部分节点可能存在性能差、安全性差等情况[27].通过入侵区块链网络中安全较为薄弱的节点,节点上部署过的所有智能合约存在被攻击者非法窃取的可能.智能合约本身的泄露对借贷等需要合约隐私性的应用场景来说是致命的,攻击者可以利用获取的智能合约分析合约的关联账户,进而推测出现实生活中与合约内容相关的用户行为,与现实生活中的用户进行身份绑定.此外,攻击者可以利用获取的智能合约进行漏洞分析,进而利用漏洞对整个系统进行攻击和系统隐私数据的窃取,其结果可能导致灾难性的系统安全问题及数据隐私泄露.

2.3 网络隐私及威胁

2.3.1 针对网络通信的隐私及威胁

相较于比特币项目提供的假匿名性,以目前市值最大的 Zcash 和 Monero[42] 为代表的加密货币系统在设计之初,通过使用 ZK-SNARK 和环签名等先进的密码学技术实现了数据隐私保护核心功能[18].尽管 Zcash 中应用的 ZK-SNARK 技术为交易信息机密性提供了良好保护、环签名技术为交易用户的身份匿名性提供了良好支持,但在远程环境下利用远程旁道攻击仍然可以窥探交易内容及用户身份信息,破坏交易机密性、不可追溯性、不可链接性及用户匿名性.攻击者利用在实现不同系统组件时泄露的旁道信息,可以将用户的所有交易链接起来,破坏交易的不可链接性,如对交易发送方和接收方的流量进行分析实施计时攻击,通过监听证明者与钱包之间的流量信息利用生成证明的时间间隔判断交易金额,破坏交易机密性;通过交易和钱包收汇款人地址之间的验证时间间隔,推断出交易与钱包地址之间的关联从而破坏交易的不可追溯性;将用户 IP 与用户的 P2P 节点链接起来从而推断用户的实际物理位置破坏用户匿名性.

2.3.2 针对网络节点的隐私及威胁

由于任意节点均可接入比特币系统,因此攻击者可以利用扫描技术、拓扑结构探测等信息收集技术对比特币网络进行探测攻击.2013 年,Reid 等人[33] 首次在小范围内通过模拟和分析比特币系统中用户的使用情况,对比特币项目隐私安全进行了较为全面的分析,实验结果表明:即使采用比特币建议的隐私保护方法,用户的隐私也无法得到保证,攻击者利用探测攻击可大量扫描区块链系统中的节点,获取节点的 IP 地址信息并分析网络规模,通过主动和被动监听的方式绘制整个系统的网络拓扑[31].Bitnodes[43] 通过在大范围内部署探测节点获取其他比特币节点信息,进而绘制出整个比特币系统的网络拓扑,甚至暴露节点的物理位置信息.当系统的网络拓扑与一些溯源技术相结合时就会严重危害用户的数据隐私安全.

此外,由于网络中的节点通常是个人电脑,在性能和安全性方面较为薄弱,因此容易受到攻击者的攻击和入侵.区块链网络是对等网络,物理位置较为分散,因此想要对整个网络采取相同的安全措施较为困难.由于区块链对等网络中的数据冗余性,入侵者可以找到一些安全较为薄弱的节点实施入侵,造成数据隐私的泄露[31].

2.4 跨链数据隐私及威胁

区块链应用落地过程中正在形成新的“数据孤岛”,跨链技术作为未来的发展方向,能够实现不同链之间的价值传递.联盟链及其落地应用更需要跨链技术来突破“围墙花园”,在不同联盟链中进行跨链操作时通常需要保证数据及账户地址的隐私性,而利用跨链技术在同构链或异构链之间进行数据交换时,跨链数据往往面临数据隐私泄露的问题,不同链之间的系统架构及实现数据隐私保护的方式可能存在差异,这将导致跨链数据隐私保护面临进一步的挑战.

在保障跨链数据隐私的前提下,各参与跨链的区块链系统需要保证其系统的安全性和数据隐私性,任何一方的系统安全性或数据隐私性存在问题都将导致跨链数据及相关账户地址的隐私泄露.此外,一些跨链技术在实现的过程中可能需要跨链桥,例如

Raze 网络[44]在实现不同链之间跨链隐私时需要 Raze 桥作为数据交换时的中间件,攻击者通过部署监听节点收集跨链桥上的通信数据,可能进一步分析出传输的具体内容,跨链桥的安全性及数据隐私性将直接影响到跨链过程中能否保障数据的隐私性.

2.5 区块链应用隐私及威胁

区块链技术在实际应用场景中存在隐私泄露问题,这类隐私泄露问题通常并非来源于区块链技术本身,而是由其他涉及区块链应用隐私泄露的主体造成,主要为区块链系统的用户及服务商.

许多用户的隐私及安全观念较为淡薄,在系统权限申请时很有可能会采取默认同意授权的方式,这为攻击者获取用户与区块链之间的交互数据提供了可乘之机.第三方对用户数据进行收集分析的情况并不少见,Englehardt 等人[45]使用开源隐私测量工具 OpenWPM1[46]对 100 多万个网站进行了测试,实验结果表明以购物网站为代表的许多网站上充斥着第三方数据隐私追踪工具.追踪者可以在匿名货币交易时获取支付流等敏感信息,如商品标识及价格等信息,通过收集足够的购买信息并将其与区块链上的交易形成一对一的对应关系,再与用户的 cookie 进行关联,进而绑定用户身份.攻击者可以进一步将多个交易与区块链交易信息联系起来,从而识别用户在整个区块链网络中的交易群及地址群.此外,用户自身对账户地址和密钥的保存方式不当也将导致严重的隐私问题,例如在论坛公开自身账户地址,攻击者可以利用论坛上用户的相关身份信息关联到实际生活中用户的身份信息,进而关联到区块链网络中的账户地址.由于用户自身密钥保存不当,攻击者获取密钥后即可获取该用户在区块链网络中所有的用户信息,造成严重的隐私泄露问题.

服务商同样存在因操作不当或服务漏洞而导致隐私泄露的可能,例如进行权限管理操作时为权限较低的用户授予了较高的权限,导致低权限用户能够获取权限之外的信息.2018 年,IOTA[47]重大盗币事件的发生为区块链应用层的隐私性及安全性敲响了警钟,此次事件造成了 1 140 万美元的损失.盗币事件发生的原因并非是 IOTA 区块链协议存在漏洞,而是攻击者利用 IOTA 为用户在线生成密钥的 Trinity 钱包插件的漏洞,通过攻击 Trinity 钱包插件不断收集用户生成的密钥种子,结合分布式拒绝服务攻击攻击(distributed denial of service, DDOS)阻止受害者收回资金,从而完成整个盗币过程.与此同时,被攻击者的身份信息及交易信息也可以被攻击者轻松获得,利用这些敏感信息,攻击者可以将区块链系统中的账户信息与现实生活中的用户身份信息进行关联,从而导致严重的隐私泄露问题.

区块链是比特币等加密电子货币的核心技术,对于实现安全、分散和开放的物联网革命至关重要.传统的集中式服务商可以非法使用物联网数据实施大规模的监控项目,如 2017 年美国国家安全局利用苹果用户的私人数据实施了棱镜项目.区块链技术结合物联网可有效避免第三方中心机构掌控用户私人数据,进而避免发生损害用户数据隐私的行为.物联网是一种新的范式,它对工作、家庭、自动化及制造等领域产生了巨大的影响,有着巨大的发展潜力[48].在物联网场景中,区块链与物联网结合尽管可以提供诸多好处,通过系统设计将隐私内置于物联网设备中,分散的物联网设备将为用户提供新的选择.在没有第三方中心机构的情况下,用户可以自己决定是否与第三方共享私人数据或出售传感器数据,私人数据将成为用户自己的财产.但在实际物联网应用场景中,由于智能物联网设备通常会在没有用户实际控制的情况下运行,因此隐私泄露问题依然严峻.

3 区块链数据隐私保护解决方案

针对区块链数据隐私泄露问题,本节从信息混淆机制、信息加密机制、通道隔离机制和权限限制机制 4 个方面详细介绍目前区块链数据隐私保护的解决方案.

信息混淆机制将交易方的地址信息、交易内容、交易方的身份信息进行混淆,解决链上交易隐私、用户身份隐私、账户地址隐私等问题;信息加密机制通过零知识证明、同态加密等加密技术对链上交易数据进行加密实现真实数据的隐藏,抵抗交易溯源、账户溯源等攻击,将加密机制与智能合约结合形成隐私智能合约可以解决交易数据隐私和跨链隐私问题;此外,信息加密机制通过洋葱路由、大蒜路由等技术抵御部署节点监听网络的攻击,实现网络信息隐藏;通道隔离机制通过闪电网络、雷电网络技术将用户间的交易移至链下进行,详细的交易信息不记录在区块链上,解决了区块链网络节点监听的问题.链上通道隔离技术将节点划分到不同的通道,每个节点只能访问所属通道内的信息,解决了节点间的隐私威胁;权限限制机制通过限制节点接入和限制信息发布解决用户端隐私、服务器端隐私和节点接入的隐私问题.

3.1 信息混淆机制

混币机制的核心思想最早源于1981年Chaum[49]发表的文章,利用第三方机构中转信息,对交易的输入输出进行混淆,从而隐藏通信双方的实际身份,实现匿名通信.混币机制是一种改变交易过程但不改变交易结果的解决方案,可归类为信息安全中的数据失真隐私保护技术.攻击者通过溯源技术可以揭露多个交易地址之间的关联性,混币机制在交易输入方及输出方之间进行交易混淆,使攻击者无法有效区分交易输入方及交易输出方,实现对交易地址之间关联性的隐藏,从而保证链上交易数据的隐私安全.在数字货币领域,根据有无第三方节点参与混币过程,可将混币机制分为中心化混币和去中心化混币2类.

3.1.1 中心化混币

中心化混币需要第三方服务商提供混币服务,混币过程由第三方服务执行.混币交易将多个交易用户的资金进行混合,将混合之后的资金输出至用户提供的输出地址,混币服务的过程中需要收取一定额度的手续费.中心化混币协议交易过程有4个阶段:

1) 协商阶段.交易用户与第三方混币服务商协定交易细节,如混币的金额、约定用户的输入地址、输出地址、混币输入输出时间、第三方混币服务商的接收地址和输出地址、混币服务手续费等.

2) 输入阶段.交易用户将约定的混币金额从输入地址发送到第三方混币服务商接收地址.

3) 输出阶段.第三方混币服务商扣除约定的手续费,然后将资产输出至交易用户的输出地址.

4) 结束阶段.混币协议正常结束后,为保护用户的隐私,所有交易用户和第三方混币服务商需要删除本次混币的协商记录.

中心化混币协议简单可行,目前已有许多第三方服务商提供混币服务.但中心化混币需要第三方节点作为中介,因此存在2个问题:

1) 信任问题.中心化混币协议缺乏对第三方混币服务商的问责机制,存在内部作恶的可能性.参与混币的用户无法确定自己的资产是否被盗刷,以及第三方混币服务提供商是否真正删除了协商记录等.

2) 第三方混币服务商提供混币服务时可能在交易时间、交易地址等方面存在规律,导致攻击者可以通过这些规律分析混币交易,进而对交易用户的输入输出地址进行关联.

Bonneau等人[50]提出了基于中心化混币的Mixcoin协议,增加了独立的加密问责机制保障用户的交易隐私,当用户隐私被窃取时,用户可以通过发布服务商的签名来降低服务商的信誉.该协议采取随机混合费用的方案避免了攻击者通过固定混合费用分析交易隐私的可能性,并将所有用户的任意组合交互放入匿名集,实现了混淆的不可区分性.Mixcoin借助匿名通信网络将多个混合交易链接在一起形成比特币混合网络从而增加了攻击难度,此方案已应用于比特币交易.

Mixcoin存在内部作恶的可能,用户的输入输出地址之间的映射对于混合服务提供商是可见的.Valenta等人[51]在Blindcoin中修改了Mixcoin协议,利用盲签名和公共日志技术保证用户输入输出地址间的映射对于混合服务提供商是不可见的.即使遭受恶意攻击,Blindcoin协议也能保证用户的强匿名性,且该协议可以抵御拒绝服务攻击.目前,该协议已应用于比特币交易.

3.1.2 去中心化混币

中心化混币通过第三方混币服务商实现交易信息的隐藏,但存在第三方泄露隐私的风险.为了彻底解决潜在的风险,Maxwell[52]最早在比特币论坛提出了基于去中心化思想的混币机制,利用用户约定的多方协议实现混币,从根本上解决了中心化混币的潜在风险.去中心化混币的交易流程分为4个阶段:

1) 协商阶段.参与混币的用户共同协商混币协议的参数.主要参数包括各个用户混币的金额、混币的输入地址、输出地址等.

2) 混淆阶段.根据去中心化混币协议对所有协商的输出地址进行混淆,同时隐藏用户的输入输出地址间的联系,保证用户的匿名性.

3) 确认阶段.经过混淆阶段得到交易的输出地址,根据交易输出地址进行交易,确认输出地址和交易无误后广播交易信息,将资产发送到对应的输出地址.

4) 结束阶段.若混合过程无错误信息,且去中心化混币协议正常结束,则所有参与此次混合的用户删除此次交易记录,混合交易结束;若混合过程出现错误,去中心化混币协议出现异常,则需要所有参与用户找出并排除恶意用户,剩余的用户重新进行混合交易.

去中心化混币不需要第三方的参与,节省了服务费用.但是去中心化混币易受到恶意攻击,通常需要对去中心化混币协议进行改进以增强其安全性,解决方案主要分为2类:

1) 改进虚拟货币协议

2013 年,Maxwell[52]提出了基于去中心化思想的 Coinjoin 协议,该协议将来自各个参与者的多个交易合并为一个交易,通过隐藏交易输入方及输出方之间的对应关系,构造混币交易,使攻击者无法通过交易信息得到输入和输出之间的关系[31].该协议虽然提高了用户的隐私保护能力,但对于行为不端的用户缺乏弹性.如果参与混合的单个用户在确认阶段拒绝签名,则整个交易就会失败,因此 Coinjoin 协议易遭受 DOS 攻击.

2014 年,Ruffing 等人[53]在 Coinjoin 协议基础上提出了一个完全去中心化的混合协议:CoinShuffle 协议,该协议规定至少有 2 个诚实的参与者,允许用户以完全匿名的方式使用比特币及其他数字货币.首先,每个参与者需要生成一个临时的加解密密钥对并对公钥进行广播;随后,参与者按照一定顺序进行排列,每个用户使用其他所有参与者的公钥对自己的输出地址进行多层加密得到新的输出地址,然后将新的输出地址发送给下一用户,混合过程保障了数据的隐私性.CoinShuffle 通过问责制确保交易完成.如果协议失败,则进入责备阶段.

2017 年,Ruffling 等人[54]在 Coinjoin 和 CoinShuffle++协议的基础上提出了 ValueShuffle 协议.该协议集成了保密交易、隐形地址和混合技术实现了全面隐私保护,包括付款人匿名、收款人匿名和支付价值隐藏,任何攻击者都不能链接到事务的输入和输出.ValueShuffle 可以在同一笔交易中混合不同价值的资金,Dos 攻击只能延迟但无法阻止协议的执行.为了避免攻击者将事务的输入与网络级标识符联系起来,建议使用外部匿名通信方式.

2) 引入新的虚拟货币

2013 年,Miers 等人[13]在比特币中加入了去中心化的加密匿名电子现金协议 Zerocoin,通过解除交易与支付来源的链接关系解决隐私泄露问题.基于数字承诺新型架构的 Zerocoin 利用零知识证明实现了强匿名性,有效防止了双重支付攻击,但 Zerocoin 实现隐私所需的成本较高.Garman 等人[55]在 Rational Zero 中改进了 Zerocoin 协议,加强 Zerocoin 匿名性的同时降低了零知识证明的资源消耗,并且确保伪造一个零币比挖矿的代价更大.但在 Zerocoin 交易过程中,会显示交易的目的地址和交易金额.Ben-Sasson[16]等利用 ZK-SNARKs 的最新技术构建了一个基于分类账且隐私保护能力更强的数字货币——Zerocash.该协议制定并构建了分散匿名支付方案,隐藏交易的支付来源、目的地址和交易金额,用户可以直接进行私下交易;Zerocash 交易效率比匿名性较低的 Zerocoin 高出多个数量级,具有较强的可用性.

3.2 信息加密机制

3.2.1 链上信息隐藏

传统的区块链交易过程中,交易信息被记录在账本中,攻击者通过账本信息获取用户账户地址、交易的金额等隐私信息.为了隐藏交易过程中相关信息,研究者们提出了同态加密、差分隐私、零知识证明、环签名、承诺方案、安全多方计算、基于属性的加密、可信硬件执行环境等技术实现数据隐私保护.安全多方计算的输入隐私特性和隐私数据联合计算功能结合隐私智能合约,保证用户输入数据的隐私.

1) 零知识证明(zero-knowledge proof)

20 世纪 80 年代,Goldwasser 等人[56]提出了零知识证明,其核心思想是证明者在不需要向验证者提供任何额外信息的前提下,使验证者相信某个论断是正确的.零知识证明分为交互式零知识证明(又称“基础零知识证明”)和非交互式零知识证明.交互式零知识证明通过证明者和验证者之间不断论述,且在论述过程中不提供任何与隐私相关信息的前提下,使验证者相信证明者论断的正确性.但是交互式零知识证明无法对证明方和验证方提前串通的行为作出判断,需要额外工作使可信第三方信服.非交互式零知识证明通过机器或程序进行试验且试验序列是保密的,不需要证明者与验证者之间的交互,完全隐藏了账本中有关交易的信息,只记录交易的存在性,解决了交互式零知识证明存在的隐患.

在区块链应用中,Zerocash 和 Zerocoin 分别通过 ZK-SNARKS 技术、零知识证明技术隐藏了交易过程中的账本信息.2016 年 Kosba 等人[57]提出了一种去中心化且能保证隐私性的智能合约框架 Hawk,区块链系统不记录交易的明文信息,以此保护交易数据的隐私安全.Hawk 的编译器可以对未加密的智能合约自动生成高效的密码协议,交易参与者使用零知识证明等密码原语与区块链进行交互,确保智能合约正确执行.Hawk 不仅能够保证链上隐私,而且能够保证同一智能合约中各方彼此之间的隐私.

2020 年,基于零知识证明的以太坊智能合约隐私协议 Zether 被 Bünz 等人[38]提出.该协议以 Zether 智能合约(zether smart contract, ZSC)的形式部署在以太坊区块链上,交易过程中的交易地址、账户

余额和交易信息始终保持加密状态从而确保数据隐私安全,实现了匿名支付.学者们利用 Bulletproofs 和 Σ 协议的特性提出了新的零知识证明机制:Σ-Bullets,并且创建了隐私账户体系,进一步增加了协议的安全性和可用性.Zether 协议主要应用场景包括保密支付、私密权益证明、保密支付、保密权益投票、保密竞拍等.Zether 隐私协议存在成本高、网络状况对交易结果影响较大等问题.类似 Zether 的其他隐私协议包括 AZTEC[58],Flyclient[59],PGC[60]等.

2) 同态加密(homomorphic encryption)

20 世纪 70 年代,Rivest 等人[61]首次提出同态加密的概念,数据经过同态加密之后再进行运算得到的结果解密后,与对明文直接进行同样运算得到的结果一致.同态加密技术隐藏了真实的账本信息,具有较高的安全性.即使同态加密结果被成功解密,攻击者仍无法知道加密前具体的数据信息,提高了信息的安全性和隐蔽性.

在区块链应用中,同态加密技术与智能合约的结合能够实现数据隐私保护.交易数据经过同态加密后发送至智能合约,智能合约对加密之后的数据进行处理,且只记录加密后的数据.攻击者无法得知加密前的数据,保证了账本信息的隐藏.2009 年,Gentry[26]提出了全同态加密算法,加密后的数据可以通过全同态加密算法可以得到与原始数据的任意运算规则相对应的运算规则,从而保证数据的同态性.但是全同态加密算法效率低,无法被大规模的使用.2011 年,Brakerski 等人[62]对 Gentry 提出的全同态加密算法进行改进,提出了基于错误学习和环错误学习假设的同态加密算法:BGV 算法.为了降低解密的复杂度、减小密文的尺寸,BGV 算法使用了新的模型转换和维数约化方法,提高了全同态加密的效率.为了突破全同态加密密钥计算的瓶颈,Gentry 等人[63]在 2013 年提出了 GSW13 算法,设计了基于属性和身份的全同态加密方案.GSW13 算法的加密密钥由每个用户的公钥组成,用户生成自己的公钥和私钥,不需要额外资源进行密钥计算从而提高了全同态加密的效率,得到了广泛的应用.

然而,利用公钥全同态加密技术实现区块链中的数据隐私保护,计算开销与通信开销巨大.为了解决公钥全同态加密开销大的问题,Zhou 和 Cao 等人[64]提出了轻量级的全同态映射数据封装机制,实现了高效的密态计算.该机制通过减少公钥加密、解密的使用次数,不依赖公钥全同态加密.该机制,利用离线状态下常数次单向陷门置换运算加密对称密钥,在线状态下仅包含乘法运算、简单加法的带密钥的对称全同态映射加密数据本身,依据混合加密基本原则,实现了“一次加密、多次使用”的目标.用户的数据集的大小不会对公钥加密的使用次数复杂度产生影响,该机制的公钥加密的使用次数复杂度为 $O(1)$.针对多用户场景,Zhou 和 Cao 等人[65]进一步构建了多密钥全同态映射数据封装机制,利用重加密的思想,在合作不合谋的双服务器模型下将不同密钥加密下的密文数据转换成统一密钥加密下的密文数据,从而提出了高效的密态计算协议,可以为区块链系统中实现加密账本与密文交易数据的统计分析及各类云计算和边缘计算中的轻量级数据隐私保护应用提供有力工具.

3) 安全多方计算(secure multi-party computation, MPC)

1982 年,为了解决百万富翁问题,姚期智基于混淆电路思想提出了两方安全计算[66].为了保护隐私安全,密码学家们经过不断研究提出了安全多方计算技术.2004 年,Malkhi 等人[67]通过支持安全两方计算的 Fairplay 系统证明了安全多方计算的可行性.2008 年,Ben-David 等人[68]提出了支持安全多方计算的 FairplayMP 系统,解决了多个互不信任的用户进行协同计算的难题.在去中心化系统中,安全多方计算能够保护输入隐私,在保证数据隐私的前提下可以进行函数计算并保证计算的准确性.

安全多方计算具有去中心化、输入隐私保护以及互不信任个体之间协同计算的特性,与区块链系统具有天然互补性.安全多方计算可以隐蔽地进行隐私数据的联合计算,将其应用于智能合约可以保证运算数据的隐私安全,实现隐私智能合约.目前已经有许多项目利用安全多方计算实现隐私智能合约保护数据的隐私安全.2015 年,Zyskind 等人[69]提出了基于安全多方计算的 Enigma 计算模型,实现了多方同时存储数据、对数据进行运算并且保持数据的完全私有.Enigma 的隐私智能合约一部分运行在公链上,用于存储智能合约正确执行的证明,另一部分运行在去中心化的链外隐私计算网络,用于执行安全多方计算.其中 MIT Enigma 应用了安全多方计算和智能合约实现了医疗数据的隐私保护.2019 年,朱岩等人[70]提出了一个基于安全多方计算的智能合约框架,利用安全多方计算、SMPC 算法和非阻塞信息传递接口技术保证了智能合约执行过程中输入隐私性和计算的正确性,并提供了强容错机制,在计算节点错误的情况下仍然可以保证群组安全通信,

确保了区块链系统中的数据隐私安全.值得注意的是,目前的安全多方计算模型的效率较低,且需要保证输入数据的真实性.

4) 承诺方案(commitment scheme)

承诺方案[71]是密码学领域中一种重要的基本协议,密码学领域中的零知识证明吸取了承诺方案的思想.此外,其对安全多方计算等加密技术也有着十分重要的地位.承诺方案是一个涉及两方的两阶段交互式协议,涉及的双方分别指承诺方和接收方.第1阶段为承诺阶段,承诺方选择一个消息 m,以密文的形式发送给接收方,意味着承诺方不会更改 m.第2阶段为打开阶段,承诺方公开消息 m 与盲化因子(相当于秘钥),接收方以此来验证其与承诺阶段所接收的消息是否一致.承诺方案有2个基本性质:隐藏性和绑定性.隐藏性指承诺值不会泄露任何关于消息 m 的信息;绑定性指任何恶意的承诺方若将承诺如果打开为非 m 的消息则不予验证通过,即接收方可以确信 m 是和该承诺对应的消息.承诺方案在分布式计算领域有广泛应用,在区块链领域可用于隐藏交易过程中的交易内容,同时可将隐藏的内容与内容所有者绑定,目前已被应用于 Zerocoin 和 Zcash 协议.

5) 环签名(ring signature)

2001年,Rivest 等人[72]提出环签名的概念.在群签名的基础上进行简化后衍生出了环签名技术,环签名在签名过程中不需要成员之间进行合作,也不需要管理者,签名者使用自身私钥和公钥池中的部分公钥即可独立进行签名.环签名具有无条件匿名性,其签名过程能够保证攻击者哪怕得到所有的公私钥后也无法得知真正的签名者是谁,同时环中成员无法伪造其他成员的签名,保证了匿名性和安全性.梁秀波等人[73]在2016年发布了一种基于环签名的区块链匿名交易方法,解决了交易双方地址暴露的问题,实现了匿名交易功能.QURAS 区块链系统通过结合环签名技术和零知识证明技术实现用户隐私保护,环签名技术使用户无法获取签名的目的地址和发送者,隐藏了用户的身份信息[63].

6) 差分隐私(differential privacy)

差分隐私[74]用来防止差分攻击,差分攻击指攻击者通过样本数据发生的新变化进行数据分析以此推断出样本数据的某些详细信息.差分隐私通过在数据中引入一定量的随机噪声,在对数据集合整体进行分析时非常接近真实结果,但攻击者无法对任何个体信息进行推断.在区块链领域,差分隐私可用于抵抗针对用户行为特征分析的攻击,在保证区块链数据整体特征保持不变的前提下,攻击者无法推测出个体用户的信息.差分隐私无法使数据完全匿名化,但可以对抗数据分析类攻击算法.

7) 基于属性加密(attribute-based encryption, ABE)

基于属性加密技术[75]有2个基本概念:属性和策略,属性指与个体相关的信息集合,策略指属性及它们之间关系所组成的逻辑表达式.合理策略后能够实现对加密数据的细粒度访问控制.此外,基于属性加密可以解决对称加密密钥传输带来的密钥泄露问题,保护了数据拥有者和数据使用者的信息.目前,基于属性加密的数据隐私保护系统已经有多个实现应用.由于基于属性加密仅从"信道安全"上部分解决了密文访问控制问题,因此从实际应用安全角度出发,Cao 等人[76]提出了"信道安全+X"安全模型,构建了可追踪、可撤销、多机构的基于属性加密方案,真正实现了高效的密文访问控制.具体而言,在可追踪方面,Liu 和 Cao 等人[77]提出了黑盒可追踪 CP-ABE 系统,首次同时达到了适应性安全和高表达力,而且该系统获得可追踪性的代价达到了目前的最佳水平.在多机构属性基加密方面,Lin 和 Cao 等人[78]提出了无中央机构的门限多机构 ABE,弱化单一属性机构被攻击带来的威胁,解决了 Sahai 和 Waters[79]在欧洲密码会议 EUROCRYPT 2005 上提出的一个公开问题.该成果被著名密码学家 Waters 在其论文[80]中作为基础文献引用,并由此提出一个新概念"Decentralizing ABE",发表在欧洲密码会议 EUROCRYPT 2011 上.但 Waters 的这项工作仍然具有"每个机构都能独立解开部分密文"的弱点.同在2011年,给出了一个标准模型下适应性安全的多机构密文策略属性基加密方案[81].在该方案中,任何机构都不能独立的解开任何密文,降低了对机构的可信性依赖,解决了属性基加密系统中固有的密钥托管问题.此外,Zhou 和 Cao 等人[82]实现了加密电子病历的多级隐私保护,给出了同时满足白盒可追踪和可撤销性质的多机构属性基加密方案.

近年来,Cao 等人[83]提出了针对代理路径的高效细粒度密文访问控制方法,即自治路径代理重加密算法.在自治路径代理重加密中,被代理者被赋予一定的优先级,按照优先级分配任务,且代理者可以自定义代理过程和路径.

8) 可信硬件执行环境

安全多方计算等复杂的密码学技术会导致区块

链系统性能降低,可信执行环境(trusted execution environment, TEE)适用于高安全需求操作、保护敏感数据和保护高价值数据等场景[84],为解决区块链性能瓶颈提供了解决方案.张凡等人[85]提出了一种经过认证的数据馈送系统 TC(town crier),TC 结合了以太坊智能合约和可信硬件,保证了与智能合约相关的数据隐私安全.由于公链的共识机制属于随机性协议,无法保证节点间信息传输一致,导致智能合约易受到回滚攻击,仅依赖 TEE 无法完全保证公链系统的数据隐私安全,因此 TEE 更适应于共识机制较为确定的联盟链.2019 年 10 月,Hyperledger Fabric 发布的 Avalon 项目结合了可信执行环境、安全多方计算和零知识证明等技术在保证数据正确性的前提下实现了数据隐私保护[86].TEE 与区块链的结合可以更好地满足区块链系统中对数据安全性和隐私性的要求,并且提供尽可能高的执行效率,提高系统可用性.

3.2.2 网络信息隐藏

传统区块链节点之间的交易信息通过假名传播,保证了一定的匿名性,但在遭受去匿名化攻击时假名仍存在泄露的风险.攻击者通过假名与 IP 地址的关联关系可以推测出 IP 地址对应的真实用户,进而对区块链节点进行审查、流量分析、服务阻断等,影响交易的正常进行.因此,网络层信息的匿名性和隐蔽性至关重要.区块链基于 P2P 网络进行消息的广播,恶意攻击者通过监听区块链的网络广播信息并将 IP 地址和链上交易信息进行链接.为防止此类攻击,需要对区块链的网络层数据进行隐藏.目前有 5 种解决方案:

1) 可信第三方转发

区块链节点间的交易需要发送者和接收者的地址信息.区块链网络层中信息的匿名传输需要保证接收方无法得知发送方的相关信息,简单可行的方法是利用可信第三方代替发送者发送信息,即发送方将信息发给可信第三方,再由可信第三方转发至接收方.目前可行的技术包括代理、虚拟私人网络等,但这类解决方案不能保证可信第三方不泄露信息,无法保证交易匿名性.

2) 混合网络

1981 年,Chaum[49]发明了混合网络.混合网络不仅混合线路节点,而且混合来自不同节点的信息.混合网络可以承受互联网服务提供商(Internet service provider, ISP)的流量分析攻击,保证了强匿名性.但混合网络延迟高,无法被大规模使用.

3) 洋葱路由

受混合网络的启发,1998 年 Paul 等人[87]发明了洋葱路由.洋葱路由通过使攻击者无法获得全局信息来保证隐私安全.为了解决高延迟的问题,洋葱路由在安全性方面做出了妥协.2004 年上线了基于洋葱路由和中继覆盖网络的强大匿名工具—Tor 网络,Tor 节点是指使用洋葱服务的洋葱路由器,能够隐藏请求端和响应端的 IP 信息,以此保证了匿名性.Tor 采用伸缩策略建立通信线路,具体流程分为 3 步.

① 节点选择阶段.请求端向目录服务器发起请求,然后通过目录服务器中的节点信息随机选择 3 个可用节点分别作为线路的入口节点、线路的中继节点和线路的出口节点.

② 线路建立阶段.请求端与入口节点建立 TLS/LLS 类型的 TCP 链接,并通过椭圆曲线密钥交换协议交换密钥;入口节点与中继节点建立 TLS/LLS 类型的 TCP 链接,请求端通过入口节点与中继节点交换密钥;最后,中继节点与出口节点建立 TLS/LLS 类型的 TCP 链接,请求端通过入口节点和中继节点与出口节点进行密钥交换.线路建立完成之后,请求端便可通过线路传输交易信息.

③ 信息传输阶段.首先,请求端需要对发送的信息进行 3 层加密,最内层的密文用出口节点密钥加密,最外层的密文用入口节点密钥加密,中间层的密文用中继节点密钥加密.然后,对信息进行传输,确保线路上的每个节点只能解开属于自己密钥加密的密文.最后,出口节点解密信息后发送给响应端.如果响应端有返回信息,则会按原路返回.

为了增强区块链系统的隐私保护,有研究者建议将 Tor 网络集成在区块链服务中.以太坊研究会也提议使用洋葱路由改进轻节点的资料可得性等.但 Tor 网络也存在一定的安全问题,攻击者通过监控多个 ISP 流量可以获取链路的组成,从而找到真正的请求端和响应端,无法确保隐私安全;如果目录服务器被破坏,洋葱路由将无法继续工作.

4) 第二代洋葱路由

Dingledine 等人[88]提出了异步、松散联合、基于电路低延迟匿名通信服务的第二代洋葱路由协议.它提供了拥塞控制服务、可配置的退出策略、目录服务器、前向保密性服务、位置隐藏服务、更好地平衡服务、完整性检查服务等,使用增量式或伸缩式路径构建方法,不再使用单一的多重信息加密.第二代洋葱路由具有良好的可用性.

5) 大蒜路由

为了解决洋葱路由存在的ISP流量攻击和目录服务器被破坏的问题,Freedman[89]提出大蒜路由,该协议具有分层加密、捆绑多份信息文件、使用EIGamal/AES加密3个主要特性.大蒜路由的典型应用是隐身互联网工程(invisible internet project, I2P),通过KAD算法获取网络节点信息生成网络数据库,由EIGamal/AES发布网络数据库条目,无需目录服务器.I2P网络使用分层加密实现隧道构建和节点路由,保证隧道中每一跳节点只知道自身相邻节点的信息.原始信息通过I2P网络被拆分为多个加密数据包,数据包上传下载相互独立的多条隧道交叉疏散传输.通过捆绑多份信息文件确定端到端间信息传递是否成功.ISP流量攻击难以追踪真实的发送端IP和接收端IP,降低了流量分析攻击成功的可能性.基于大蒜路由的I2P网络主要工作流程有2个步骤.

① 隧道构建阶段.由于隧道是单向的,因此发送端和接收端需要分别建立属于自己的输入输出隧道.通信隧道的默认长度为3跳,完成一次完整的通信需要4条隧道[90].

② 数据加密和传输阶段.首先待处理的信息由发送端进行3次层叠式加密,将处理得到的密文传输到发送端通信输出隧道的Gateway节点,其次需要在隧道中各个节点上依次解密,然后将解密后的结果转发到接收端输入隧道的Gateway节点[90].由接收端Gateway节点加密转发到接收端,最后接收端经过解密得到信息.从接收端到发送端的消息传输原理一致.

3.3 通道隔离机制

为保证网络层的数据隐私安全,研究者提出了区块链通道机制实现账本数据的隔离.非许可链中需要实现链下通道隔离,即将区块链上的交易移至链下进行.对于许可链来说,通道隔离主要应用于联盟链中,不同联盟成员属于不同通道,通道隔离技术使数据仅对通道内的节点可见.

3.3.1 链下通道隔离

为了减少区块链上的交易压力、提高运行效率,状态通道技术应运而生.状态通道的核心思想是将区块链上的一些交易移到链下进行,经典的状态通道技术有闪电网络和雷电网络2种,目前主要应用于资产交易.随着区块链技术的不断发展,人们对隐私的要求不断提高,研究人员将闪电网络和雷电网络用于保护交易过程中用户的隐私安全.闪电网络技术主要应用于比特币,雷电网络技术主要应用于以太坊.

1) 闪电网络技术

传统比特币交易需要较高的交易费用和较长的交易时间,不适用于高频小额支付,且易造成用户交易信息泄露.为了实现高频小额支付、提高用户交易信息隐私安全性,2015年闪电网络的概念由Poon和Dryja[91]首次提出,并在2016年发表了对闪电网络详细描述的论文.

闪电网络运行在区块链的链下,通过链下通道实现高频小额支付,交易费用低,交易信息不记录在区块链上.用户间使用闪电网络进行链下交易时,首先在区块链上新建交易,然后打开双向支付通道,交易双方记录通道上的账本,发生资产变化时及时更新账本;交易双方的资产被保存在多重签名钱包中,多重签名钱包指多人共同管理同一笔资金的钱包,交易的一组私钥可以查看交易双方的资金.当交易双方对最终交易签名或者交易到达了设置时间就结束交易,交易结束后多重签名钱包会将资金返还给交易双方;根据交易双方对于资产划分的规则生成承诺交易并进行签名,通过承诺交易进行资产转移,然后更新支付通道账本中各自的余额信息;链下交易结束后,交易双方需要用私钥签署交易,将双方余额发送到区块链上.此过程属于单通道支付,交易双方直接通过通道交易.如果交易双方A和B不能直接建立支付通道,但存在中间方C,且A和C、B和C之间都存在支付通道,则A和B可以通过C进行交易.闪电网络通过哈希时间锁定合约(hash time lock contract, HTLC)保证了交易的安全性.

闪电网络技术的链下交易方案实现了高频小额交易.链下交易不仅保证了用户的隐私,而且减少了区块链的交易压力,增强了区块链的可扩展性.

2) 雷电网络技术

雷电网络主要应用于以太坊,其通道技术以及链下交易流程与闪电网络类似.雷电网络基于智能合约制定了链下通道中交易双方共识的资产划分规则.链下交易通道关闭时,由智能合约对交易进行清算并将交易双方的余额发到区块链上.雷电网络以托管形式持有代币,用于链下交易,不需要多重签名钱包.雷电网络在链下交易过程中实现了智能转账.闪电网络中的HTLC可能存在A不想通过C向B转账的情况,若A想改变路径与B交易,则需要等待HTLC到期,A不能单方面结束交易.在这个时间间隙,B和C可能会串通导致A在原来路径中的资金遭到损失.雷电网络利用重试哈希锁、收据哈希

锁和时间锁构成组合锁,保证在多支付通道中交易双方的资产不受损失.

雷电网络技术相对于闪电网络技术来说,通过智能合约为交易双方制定交易规则,使用代币用于链下交易,实现了智能转账,利用组合锁保证用户的资金不受损失,交易的隐私性和安全性得到保证.

3.3.2 多链通道隔离

实际应用场景中,存在多个组织之间用户进行交互的情况,每个组织内部的信息需要通过通道机制实现隐私保护.多链通道隔离技术将一个区块链系统内的节点划分到多个通道,每个通道内维护1个子账本,通道与通道之间相互隔离,通过身份认证、访问控制机制保护通道内节点的数据隐私.多通道技术首先要严格进行节点身份管理,对通道内的节点进行授权,节点加入通道时进行身份认证,不同身份的节点其权限可能不同.其次,多通道技术要严格进行通道管理,主要体现在通道的建立、运行维护和销毁等.最后,多通道技术要严格进行事务管理.

2017年 Hyperledger Fabric 提出的通道机制是区块链中多链通道技术最成熟的应用.Fabric 的主要应用场景是联盟链,各个联盟内部形成一个通道,通过成员身份管理维护通道内的账本,Fabric 网络中的通道隔离机制如图3所示.假设一个 Fabric 联盟网络 N 有4个组织 R_1,R_2,R_3,R_4,组织对应的证书颁发机构为 CA_1,CA_2,CA_3,CA_4.组织 R_1 与 R_4 根据网络配置文件 NC_4 的内容对联盟网络 N 进行管理.通道 C_1,C_2 的通道配置文件分别是 CC_1,CC_2,通道 C_1 由组织 R_1 与 R_2 管理,通道 C_2 由组织 R_2 与 R_3 管理.建立网络时,证书颁发机构 CA_4 先对组织 R_4 授权,组织 R_4 的排序节点 O_4 作为网络的初始化管理者,有权设置网络的初始版本.组织 R_1 与 R_2、R_2 与 R_3 需要在网络中正常通信.R_1,R_2,R_3 的客户端应用 A_1,A_2,A_3 分别在通道 C_1,C_1 与 C_2,C_3 中进行业务交易.节点 P_1 维护 C_1 通道内的账本信息,节点 P_2 同时维护 C_1,C_2 通道的账本信息,节点 P_3 维护通道 C_2 的账本信息.如果通道 C_1,C_2 之间需要通信,则需要通过节点 P_2 实现2个通道间的交易.

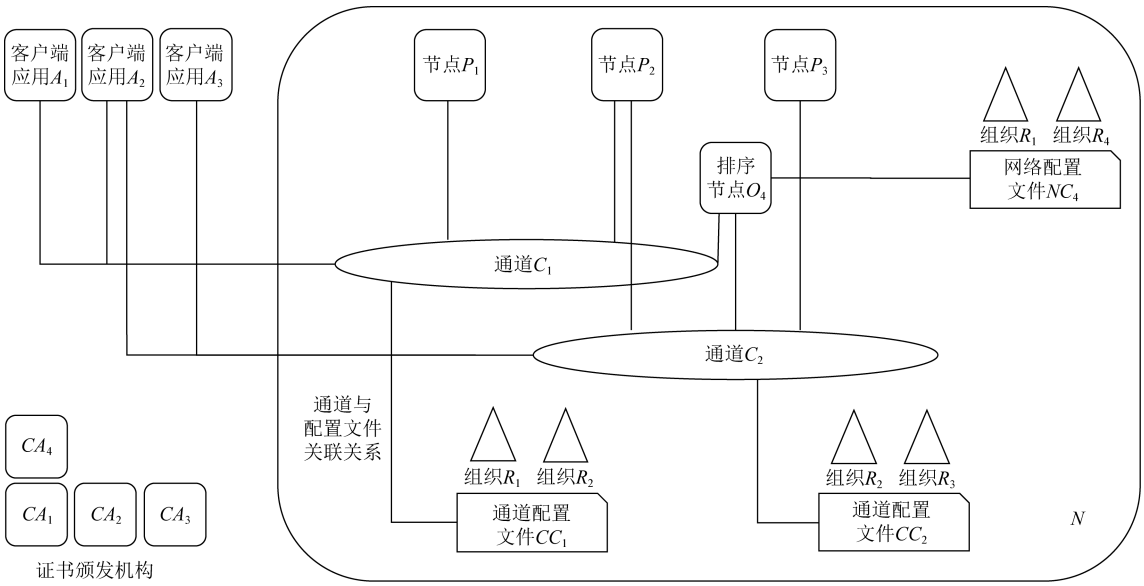

Fig. 3 Fabric muti-channel network[92]

图3 Fabric 多通道网络[92]

链下通道隔离主要应用于非许可链,多链通道隔离主要应用于许可链.对于非许可链,账本信息由所有节点进行维护.因此为了保护节点的数据隐私安全,将节点之间的交易移至链下进行.对于许可链,通过节点间的身份认证对节点进行分类和管理,通过通道技术保护各个通道内节点的隐私安全.

3.4 权限限制机制

如果区块链系统允许任意节点加入,则恶意节点可能会加入到区块链系统中窃取账本信息和其他节点的隐私数据等,导致区块链系统的隐私泄露.为了保护区块链的数据隐私安全,需要对节点的接入进行身份认证,限制恶意节点和不符合系统要求的节点接入区块链系统.另外,用户身份管理机制也可以保护用户的隐私安全.

3.4.1 限制节点接入

许可链指参与到区块链网络的节点都是经过

许可的,私有链和联盟链都属于许可链.非许可链对加入节点不设置限制条件,是完全去中心化的区块链系统.许可链和非许可链的身份认证机制不同,因此分为非许可链节点身份认证和许可链节点身份认证2方面介绍.

1) 非许可链节点身份认证.对于非许可链来说,节点接入不需要通过第三方认证,因此只能通过P2P网络层对节点接入进行限制.目前主要从节点行为检测和信任关系建立2方面对非许可链的节点进行身份认证[93].研究者提出了多种解决方案来识别异常节点.2013年,You等人[94]提出通过N-gram算法对节点的历史行为进行分析,识别异常节点并对节点的可用性进行评估.2016年,Epishkina等人[95]通过T模式对节点进行隐藏事件模式聚类,该模式为了找到隐藏的事件模式,要求对每个事件的重复次数大于等于2次.2017年,行为模式聚类算法由Huang等人[96]提出,其利用行为模式聚类进行恶意节点检测,与其他方法相比,该方案可以检测出区块链上较多的恶意节点.杨保华等人[56]利用共识机制检测并排除恶意节点,该机制简单易用,因此被广泛运用于一些开源的区块链项目.通过识别并且限制恶意节点加入区块链,在一定程度上保证非许可区块链中节点的数据隐私安全.

建立信任关系指想要接入区块链系统的节点与已存在于区块链系统中的节点之间构建信任关系,通过节点之间互相认证确保新加入的节点不是恶意节点.为了抵御Sybil攻击(Sybil攻击是指多节点、多身份的攻击),2006年,王鹏等人[97]提出了限制节点加入的认证方案,通过提高节点加入网络的代价,从而提高攻击者进行大规模攻击的代价.网络规模越大,对新节点验证需要的资源越多,操作越困难.2008年,Yu等人[98]提出了SybilGuard协议,减小了Sybil攻击的破坏性.SybilGuard协议建立节点之间的信任网络,恶意节点虽然可以创建很多身份,但是这些节点间的信任关系很少.通过随机游走将恶意节点与诚实节点进行分割,并过滤恶意节点从而限制恶意节点创建身份的数量.2017年,Fang等人[99]利用趋势分析和差异判断识别节点信誉值的差异,设置信任阈值和累积信誉度.如果信任度累积达到了阈值则可以建立信任关系.与其他方法相比,该方案易于实现.

2) 许可链节点身份认证.为了保证许可链上节点的隐私安全,新的节点加入区块链系统时,应避免无身份节点或恶意节点的加入.因此,需要对节点进行身份认证.许可链通过可信第三方签发的数字证书对节点进行身份认证.

Hyperledge Fabric 1.3中提出了Fabric-CA用于对入网节点的身份进行认证[92].节点加入Fabric网络需要先向CA发起请求,如果节点符合Fabric链的要求,则为节点颁发数字证书,包括发行担保证书和发行交易证书.最后节点经过身份认证接入区块链系统,避免节点的数据隐私被恶意节点泄露,保证了区块链系统的安全.远程证明是可信计算的重要功能[100],但远程证明模型是面向中心化网络设计的.2018年刘明达等人[101]将区块链技术与远程证明结合,提出了基于区块链的远程证明模型(remote attestation model based on blockchain, RABBC).节点不仅需要拥有可信平台模块(trusted platform module, TPM)标识证明身份,同时也要证明自身处于可信状态.该模型底层采用直接匿名证明(direct anonymous attestation, DAA)认证协议实现了节点身份认证并保证了匿名性.

对许可链和非许可链的节点权限控制都是为了防止恶意节点接入区块链系统,从而避免恶意攻击和节点数据隐私泄露.为了保护数据隐私安全,非许可链节点权限控制主要是通过对网络中节点的行为分析预测恶意节点,或通过节点之间的信任关系辨别出恶意节点;许可链节点权限控制主要是通过为节点颁发数字证书进行身份认证.

3.4.2 限制数据发布

为了实现区块链上的数据隐私保护,可以对上链前的数据进行筛选过滤.从源头上限制数据的发布,从而实现数据隐私保护.2020年Yong等人[102]提出了可编辑区块链框架,在保证区块链安全可信的前提下实现链上数据的可编辑操作.该框架可以对上链前的数据进行过滤,限制一些数据的发布,从而保证数据的隐私安全.

3.4.3 自我主权身份模型

隐私保护身份管理(identity management, IdM)模式加强了区块链的隐私保护,但需要第三方机构集中式对用户身份进行管理.基于自我主权身份(self-sovereign identity, SSI)[103]概念的身份管理模式能够实现SSI用户对自身所有个人信息和相关活动数据拥有完全的自主权,同时保证行驶自主权时的安全性和效率.SSI指区块链系统中的用户在任何时间、任何地点、任何在线情况下都可以自主的控制个人隐私数据.用户之外的第三方服务所拥有的的数据都不是原始的,保证了用户的隐私数据不被

第三方泄露.自我主权身份模式为用户提供了匿名机制,在用户进行交易时控制他们的隐私数据,保证隐私安全[104].SSI 概念的主要过程和相关实体如图 4 所示.用户以自己为中心使用手机从发行机构获取证书、ZK 凭证和可验证的声明,其中用户的私钥被保护在钱包中.发行人向区块链发送用户的身份哈希、证书哈希和可验证的声明,然后用户向区块链进行 SSI 身份认证.为了提高 SSI 模型的隐私保护功能,用户需要向服务提供者提供零知识加密证明,服务提供者在区块链上对用户签名、身份进行检查.

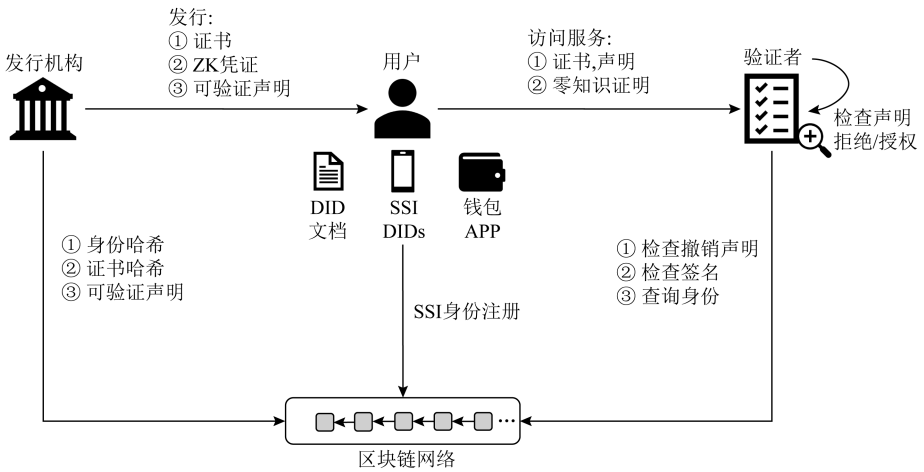

Fig. 4 Self-sovereign identity management model in blockchain[105]

图 4 区块链自我主权身份管理模型[105]

4 区块链数据隐私保护技术发展展望

4.1 量子计算

密码学安全技术为区块链技术的诞生提供了必要的前提,但密码学安全技术可能存在漏洞及未来被破解的可能性,这将对区块链系统造成毁灭性打击.1981 年费曼提出量子计算的概念,即利用量子比特可以同时处于多个相干叠加态的特性,通过并行处理来提高计算的速度[56].在量子计算条件下,1994 年提出的基于量子计算的 Shor 算法对 DES 等算法具有很大的威胁.若区块链系统所使用的密码学技术被破解,则建立于被破解密码学技术之上的区块链系统的安全性和隐私性将受到严峻挑战.因此,如何实现量子计算条件下安全可靠的密码学安全技术将直接影响区块链技术的未来发展,必将成为未来区块链技术研究的热点.

4.2 跨链数据隐私保护标准

区块链技术要实现广领域大规模应用需要成熟的跨链技术作为不同区块链平台之间的桥梁,同时需要更高效的隐私保护算法.跨链技术是实现各个区块链之间数据交互和价值传递的关键.目前,利用跨链原子交换协议实现了区块链之间的资产转移,进行跨链交易时任意交易双方的身份信息及交易内容都是公开的,对于跨链交易数据隐私保护问题的解决方案还较为稀缺,仍需要进一步研究.区块链大规模应用场景中,如电子商务[106]和智慧城市[107]等,不同应用对隐私保护的要求标准有所不同,具体实现隐私保护的技术也有所区分,导致了不同区块链系统隐私保护技术的碎片化和多样化,使得区块链系统之间彼此很难集成到一起.在实现区块链跨链隐私保护方面,W3C 正在对一些涉及隐私保护模型和技术相关的构件进行标准化,如声明验证[108]和去中心化身份识别器[109]等.目前,一些含隐私保护功能的区块链系统,如 Uport[110]和 Sovrin[111]正计划采用这些标准实现对应的隐私功能.区块链行业应尽早对数据隐私保护问题制定相应的标准,在隐私保护基准的前提下实现跨链连接.标准的制定需要整个区块链行业研究人员的共同参与,并且得到区块链行业的广泛认可,如何制定标准、制定什么样的标准是实现跨链数据隐私保护需要面对的挑战.

4.3 数据隐私保护监管标准

数字货币的匿名性极易成为洗钱组织、毒品交易组织及恐怖组织等不法分子的犯罪工具.针对反洗钱、反恐怖等社会安全需求,数字货币的颁发机构需要与监管机构相互协调,满足数字货币匿名性的同时,也能满足执法机构对犯罪分子的追踪.当区块链技术与数字货币领域结合时,相关的数据隐私信息

除了货币的使用者及政府监管部门之外,应保证其他人无法获得货币的任何相关隐私信息如货币的历史拥有者、历史参与的交易等,确保交易过程的匿名性.如何让政府监管部门理解区块链技术并适度监管是目前区块链应用面临的一个重要挑战.区块链技术本身的去中心化特性与匿名性容易将区块链的应用场景导向一些不易监管的领域,与此同时,政府部门对区块链技术的监管程度也将直接影响区块链技术的发展前景,只有在区块链技术与政府监管标准之间找到合适的平衡点才能促进区块链应用的技术发展和实际落地.例如通用数据保护条例(general data protection regulation, GDPR)[112]等规定有可能与区块链技术存在冲突,区块链本身的不可篡改性、持久性与该规定中用户拥有更改和删除自身隐私的权利相冲突.此外,2019 年中国国家互联网信息办公室室务会议审议通过了《区块链信息管理服务规定》[113],其对中国区块链信息服务提出了一系列标准和规定,开发者在区块链实际应用设计和开发过程中应遵循相应的规定.

5 总 结

随着业内关于区块链技术的认知不断深入以及各场景下区块链系统落地数量的逐步增长,数据隐私保护问题将成为专家学者的重点研究方向.本文结合区块链技术的历史发展进程,介绍了各发展阶段中的数据隐私泄露问题,提出了针对区块链技术的数据隐私定义,并根据区块链相关技术要点详细介绍了相应隐私泄露问题及数据隐私保护所面临的挑战.本文从信息混淆机制、信息加密机制、权限限制机制、通道隔离机制的角度总结了相应的解决方案,最后根据区块链数据隐私保护的研究现状,对未来区块链数据隐私保护的研究方向进行了展望.

参 考 文 献

[1] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system [EB/OL].[2021-08-15]. https://bitcoin.org/bitcoin.pdf

[2] Kano Y, Nakajima T. A novel approach to solve a mining work centralization problem in blockchain technologies [J]. International Journal of Pervasive Computing and Communications, Emerald Publishing Limited, 2018, 14(1): 15-32

[3] DeGroot M H. Reaching a consensus [J]. Journal of the American Statistical Association, Taylor & Francis Group, 1974, 69(345): 118-121

[4] Schollmeier R. A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications [C] //Proc 1st Int Conf on Peer-to-Peer Computing. Piscataway, NJ: IEEE, 2001: 101-102

[5] Buterin V. A next-generation smart contract and decentralized application platform [J]. White Paper, 2014, 3(37) [2021-08-15]. https://translatewhitepaper.com/wp-content/uploads/2021/04/EthereumOrijinal-ETH-English.pdf

[6] Diffie W, Hellman M. New directions in cryptography [J]. IEEE Transactions on Information Theory, 1976, 22(6): 644-654

[7] Swan M. Blockchain: Blueprint for a New Economy [M]. Sebastopol, CA: O'Reilly Media, 2015

[8] Schwab K, Marcus A, Oyola J O, et al. Personal data: The emergence of a new asset class [EB/OL]. [2021-08-15]. https://www.weforum.org/reports/personal-data-emergence-new-asset-class

[9] Charlie L. Litecoin-P2P digital coin [EB/OL]. [2021-05-22]. https://litecoin.org/cn/

[10] Jonathan K. QuarkChain [EB/OL]. [2021-08-15]. https://quarkchain.io/cn/

[11] Tecshare. Infinitecoin [EB/OL]. [2021-08-15]. http://infinitecoin.com/

[12] Goldreich O, Oren Y. Definitions and properties of zero-knowledge proof systems [J]. Journal of Cryptology, 1994, 7(1): 1-32

[13] Miers I, Garman C, Green M, et al. Zerocoin: Anonymous distributed E-Cash from Bitcoin [C] //Proc of 2013 IEEE Symp on Security and Privacy. Berkeley, CA: IEEE, 2013: 397-411

[14] Sasson E B, Chiesa A, Garman C, et al. Zerocash: Decentralized anonymous payments from bitcoin [C] //Proc of 2014 IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2014: 459-474

[15] Zooko W O. Privacy-protecting digital currency-Zcash [EB/OL]. [2021-04-17]. https://z.cash/technology/

[16] Ben-Sasson E, Chiesa A, Tromer E, et al. Succinct non-interactive zero knowledge for a von Neumann architecture [C] //Proc of the 23rd USENIX Conf on Security Symp. Berkeley, CA: USENIX Association, 2014: 781-796

[17] Zhang Fangguo, Kim K. ID-based blind signature and ring signature from pairings [C] //Proc of the Int Conf on the Theory and Application of Cryptology and Information Security. Berlin: Springer, 2002: 533-547

[18] Tramèr F, Boneh D, Paterson K. Remote side-channel attacks on anonymous transactions [C] //Proc of the 29th USENIX Security Symp. Berkeley, CA: USENIX Association, 2020: 2739-2756

[19] Yuan Yong, Wang Feiyue. Blockchain: The state of the art and future trends [J]. Acta Automatica Sinica, 2016, 42(4): 481-494 (in Chinese)
(袁勇, 王飞跃. 区块链技术发展现状与展望[J]. 自动化学报, 2016, 42(4): 481-494)

[20] Androulaki E, Barger A, Bortnikov V, et al. Hyperledger fabric: A distributed operating system for permissioned blockchains [C] //Proc of the 13th EuroSys Conf. New York: ACM, 2018: 1-15

[21] WeBank. FISCO BCOS blockchain [EB/OL]. [2021-08-15] https://fisco-bcos-documentation. readthedocs. io/zh _ CN/latest/docs/introduction.html.

[22] Brandeis L, Warren S. The right to privacy [J]. Harvard Law Review, 1890, 4(5): 193-220

[23] China Institute of Electronic Technology Standardization. China blockchain technology industry development forum [EB/OL]. [2021-08-15]. http://www.cbdforum.cn/bcweb/

[24] Özsu M T, Valduriez P. Principles of Distributed Database Systems [M]. Berlin: Springer, 2020

[25] Wang Huaqun, Zhang Fan, Li Tian, et al. Security and privacy protection technologies in smart contract [J]. Journal of Nanjing University of Posts and Telecommunications, 2019, 39(4): 63-71 (in Chinese)
(王化群, 张帆, 李甜, 等. 智能合约中的安全与隐私保护技术[J]. 南京邮电大学学报: 自然科学版, 2019, 39(4): 63-71)

[26] Gentry C. A Fully Homomorphic Encryption Scheme [M]. Stanford, CA: Stanford University, 2009

[27] Shor P W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer [J]. SIAM Review, 1999, 41(2): 303-332

[28] Long Guilu. Grover algorithm with zero theoretical failure rate [J]. Physical Review A, 2001, 64(2): No.022307

[29] Hwang W-Y. Quantum key distribution with high loss: Toward global secure communication [J]. Physical Review Letters, 2003, 91(5): No.057901

[30] Scarani V, Bechmann-Pasquinucci H, Cerf N J, et al. The security of practical quantum key distribution [J]. Reviews of Modern Physics, 2009, 81(3): 1301-1350

[31] Zhu Liehuang, Gao Feng, Shen Meng, et al. Survey on privacy preserving techniques for blockchain technology [J]. Journal of Computer Research and Development, 2017, 54(10): 2170-2186 (in Chinese)
(祝烈煌, 高峰, 沈蒙, 等. 区块链隐私保护研究综述[J]. 计算机研究与发展, 2017, 54(10): 2170-2186)

[32] Reid F, Harrigan M. An Analysis of Anonymity in the Bitcoin System [M]. Berlin: Springer, 2013: 197-223

[33] Androulaki E, Karame G O, Roeschlin M, et al. Evaluating User Privacy in Bitcoin [M]. Berlin: Springer, 2013: 34-51

[34] Ron D, Shamir A. Quantitative analysis of the full bitcoin transaction graph [C] //Proc of the Int Conf on Financial Cryptography and Data Security. Berlin: Springer, 2013: 6-24

[35] Meiklejohn S, Pomarole M, Jordan G, et al. A fistful of bitcoins: Characterizing payments among men with no names [C] //Proc of the 2013 Conf on Internet Measurement Conf. New York: ACM, 2013: 127-140

[36] Monaco J V. Identifying bitcoin users by transaction behavior [C/OL] //Biometric and Surveillance Technology for Human and Activity Identification XII. Bellingham, Washington: Int Society for Optics and Photonics, 2015, 9457: 945704 [2021-08-15]. https://doi.org/10.1117/12.2177039

[37] Steffen S, Bichsel B, Gersbach M, et al. Zkay: Specifying and enforcing data privacy in smart contracts [C] //Proc of the 2019 ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2019: 1759-1776

[38] Bünz B, Agrawal S, Zamani M, et al. Zether: Towards privacy in a smart contract world [C] //Proc of the Int Conf on Financial Cryptography and Data Security. Berlin: Springer, 2020: 423-443

[39] Atzei N, Bartoletti M, Cimoli T. A survey of attacks on ethereum smart contracts (sok)[C] //Proc of the Int Conf on Principles of Security and Trust. Berlin: Springer, 2017: 164-186

[40] Liu Chao, Liu Han, Cao Zhao, et al. Reguard: Finding reentrancy bugs in smart contracts [C] //Proc of 2018 IEEE/ACM 40th Int Conf on Software Engineering: Companion (ICSE-Companion). Piscataway, NJ: IEEE, 2018: 65-68

[41] Luu L, Chu D H, Olickel H, et al. Making smart contracts smarter [C] //Proc of the 2016 ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2016: 254-269

[42] Nakamoto S. Monero project [EB/OL]. [2021-05-28]. https://www.getmonero.org//index.html

[43] 21 Company. Global Bitcoin nodes distribution [EB/OL]. [2021-08-15]. https://bitnodes.io/

[44] Raze Network. Trustless Privacy on Polkadot [EB/OL]. [2021-08-15]. https://www.raze.network/

[45] Englehardt S, Narayanan A. Online tracking: A 1-million-site measurement and analysis [C] //Proc of the 2016 ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2016: 1388-1401

[46] Acar G, Eubank C, Englehardt S, et al. The web never forgets: Persistent tracking mechanisms in the wild [C] //Proc of the 2014 ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2014: 674-689

[47] IOTA Chinese Community. IOTA [EB/OL]. [2021-08-15]. https://www.iotachina.com/ (in Chinese)
(IOTA 中国社区. IOTA 埃欧塔[EB/OL]. [2021-08-15]. https://www.iotachina.com/)

[48] Atzori L, Iera A, Morabito G. The Internet of Things: A survey [J]. Computer Networks, 2010, 54(15): 2787-2805

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

[50] Bonneau J, Narayanan A, Miller A, et al. Mixcoin: Anonymity for bitcoin with accountable mixes [C] //Proc of the Int Conf on Financial Cryptography and Data Security. Berlin: Springer, 2014: 486-504

[51] Valenta L, Rowan B. Blindcoin: Blinded, accountable mixes for bitcoin [C] //Proc of the Int Conf on Financial Cryptography and Data Security. Berlin: Springer, 2015: 112-126

[52] Maxwell G. CoinJoin: Bitcoin privacy for the real world [EB/OL]. [2021-08-15]. https://bitcointalk.org/index.php?topic=279249.0

[53] Ruffing T, Moreno-Sanchez P, Kate A. Coinshuffle: Practical decentralizedcoin mixing for bitcoin [C] //Proc of the European Symp on Research in Computer Security. Berlin: Springer, 2014: 345-364

[54] Ruffing T, Moreno-Sanchez P. Valueshuffle: Mixing confidential transactions for comprehensive transaction privacy in bitcoin [C] //Proc of the Int Conf on Financial Cryptography and Data Security. Berlin: Springer, 2017: 133-154

[55] Garman C, Green M, Miers I, et al. Rational zero: Economic security for zerocoin with everlasting anonymity [C] //Proc of the Int Conf on Financial Cryptography and Data Security. Berlin: Springer, 2014: 140-155

[56] Goldwasser S, Micali S, Rackoff C. The knowledge complexity of interactive proof systems [J]. SIAM Journal on Computing, 1989, 18(1): 186-208

[57] Kosba A, Miller A, Shi E, et al. Hawk: The blockchain model of cryptography and privacy-preserving smart contracts [C] //Proc of the 2016 IEEE Symp on Security and Privacy (SP). Piscataway, NJ: IEEE, 2016: 839-858

[58] Williamson Z J. The aztec protocol [EB/OL]. 2018 [2021-08-15]. https://raw.githubusercontent.com/AztecProtocol/AZTEC/master/AZTEC.pdf

[59] Bünz B, Kiffer L, Luu L, et al. Flyclient: Super-light clients for cryptocurrencies [C] //Proc of the 2020 IEEE Symp on Security and Privacy (SP). Piscataway, NJ: IEEE, 2020: 928-946

[60] Chen Yu, Ma Xuecheng, Tang Cong, et al. PGC: Decentralized confidential payment system with auditability [C] //Proc of the European Symp on Research in Computer Security. Berlin: Springer, 2020: 591-610

[61] Rivest R L, Adleman L, Dertouzos M L. On data banks and privacy homomorphisms [J]. Foundations of Secure Computation, 1978, 4(11): 169-180

[62] Brakerski Z, Vaikuntanathan V. Efficient fully homomorphic encryption from (standard) LWE [J]. SIAM Journal on Computing, 2014, 43(2): 831-871

[63] Gentry C, Sahai A, Waters B. Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based [C] //Proc of the Annual Cryptology Conf. Berlin: Springer, 2013: 75-92

[64] Zhou Jun, Choo K K R, Cao Zhenfu, et al. PVOPM: Verifiable privacy-preserving pattern matching with efficient outsourcing in the malicious setting [J]. IEEE Transactions on Dependable and Secure Computing, 2019 [2021-08-15]. doi:10.1109/TDSC.2019.2947436

[65] Zhou Jun, Cao Zhenfu, Qin Zhan, et al. LPPA: Lightweight privacy-preserving authentication from efficient multi-key secure outsourced computation for location-based services in VANETs [J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 420-434

[66] Yao A C. Protocols for secure computations [C] //Proc of the 23rd Annual Symp on Foundations of Computer Science. Piscataway, NJ: IEEE, 1982: 160-164

[67] Malkhi D, Nisan N, Pinkas B, et al. Fairplay-secure two-party computation system [C/OL] //Proc of the USENIX Security Symp. 2004, 4: 9 [2021-08-15]. https://www.usenix.org/event/sec04/tech/full_papers/malkhi/malkhi_html/

[68] Ben-David A, Nisan N, Pinkas B. FairplayMP: A system for secure multi-party computation [C] //Proc of the 15th ACM Conf on Computer and Communications Security. Alexandria, VA: The Association for Computing Machinery, 2008: 257-266

[69] Zyskind G, Nathan O, Pentland A. Enigma: Decentralized computation platform with guaranteed privacy [J]. ArXiv Preprint, ArXiv:1506.03471, 2015

[70] Zhu Yan, Song Xiaoxun, Xue Xianbin, et al. Smart contract execution system over blockchain based on secure multi-party computation [J]. Journal of Cryptologic Research, 2018, 6(2): 246-257 (in Chinese)
(朱岩, 宋晓旭, 薛显斌, 等. 基于安全多方计算的区块链智能合约执行系统[J]. 密码学报, 2018, 6(2): 246-257)

[71] Brassard G, Chaum D, Crépeau C. Minimum disclosure proofs of knowledge [J]. Journal of Computer and System Sciences, 1988, 37(2): 156-189

[72] Rivest R L, Shamir A, Tauman Y. How to leak a secret [C] //Proc of the Int Conf on the Theory and Application of Cryptology and Information Security. Berlin: Springer, 2001: 552-565

[73] Liang Xiubo, Li Qilei, Yin Keting, et al. A kind of block chain anonymous deal method based on ring signatures: China, CN106779704A [P/OL]. (2017-05-31) [2021-08-15]. https://patents.google.com/patent/CN106779704A/zh (in Chinese)
(梁秀波, 李启雷, 尹可挺, 等. 一种基于环签名的区块链匿名交易方法: 中国, CN106779704A [P/OL]. (2017-05-31) [2021-08-15]. https://patents.google.com/patent/CN106779704A/zh)

[74] Dwork C, Roth A. The algorithmic foundations of differential privacy [J]. Found Trends Theor Computer Science, 2014, 9(3-4): 211-407

[75] Bethencourt J, Sahai A, Waters B. Ciphertext-policy attribute-based encryption [C] //Proc of the 2007 IEEE Symp on Security and Privacy (SP'07). Piscataway, NJ: IEEE, 2007: 321-334

[76] Cao Zhenfu, Dong Xiaolei, Zhou jun, et al. Research advances on big data security and privacy preserving [J]. Journal of Computer Research and Development, 2016, 53(10): 2137-2151 (in Chinese)

(曹珍富, 董晓蕾, 周俊, 等. 大数据安全与隐私保护研究进展[J]. 计算机研究与发展, 2016, 53(10): 2137-2151)

[77] Liu Zhen, Cao Zhenfu, Wong D S. Blackbox traceable CP-ABE: How to catch people leaking their keys by selling decryption devices on ebay [C] //Proc of the 2013 ACM SIGSAC Conf on Computer & Communications Security. New York: Association for Computing Machinery. 2013: 475-486

[78] Lin Huang, Cao Zhenfu, Liang Xiaohui, et al. Secure threshold multi authority attribute based encryption without a central authority [C] //Proc of the Int Conf on Cryptology in India. Berlin: Springer, 2008: 426-436

[79] Sahai A, Waters B. Fuzzy identity-based encryption [C] //Proc of the Annual Int Conf on the Theory and Applications of Cryptographic Techniques. Berlin: Springer, 2005: 457-473

[80] Lewko A, Waters B. Decentralizing attribute-based encryption [C] //Proc of the Annual Int Conf on the Theory and Applications of Cryptographic Techniques. Berlin: Springer, 2011: 568-588

[81] Liu Zhen, Cao Zhenfu, Huang Qiong, et al. Fully secure multi-authority ciphertext-policy attribute-based encryption without random oracles [C] //Proc of the European Symp on Research in Computer Security. Berlin: Springer, 2011: 278-297

[82] Zhou Jun, Cao Zhenfu, Dong Xiaolei, et al. TR-MABE: White-box traceable and revocable multi-authority attribute-based encryption and its applications to multi-level privacy-preserving e-healthcare cloud computing systems [C] //Proc of the 2015 IEEE Conf on Computer Communications (INFOCOM). Piscataway, NJ: IEEE, 2015: 2398-2406

[83] Cao Zhenfu, Wang Hhongbing, Zhao Yunlei. AP-PRE: Autonomous path proxy re-encryption and its applications [J]. IEEE Transactions on Dependable and Secure Computing, 2017, 16(5): 833-842

[84] Sabt M, Achemlal M, Bouabdallah A. Trusted execution environment: What it is, and what it is not [C] //Proc of the 2015 IEEE Trustcom/BigDataSE/ISPA. Piscataway, NJ: IEEE, 2015: 57-64

[85] Zhang Fan, Cecchetti E, Croman K, et al. Town crier: An authenticated data feed for smart contracts [C] //Proc of 2016 ACM SIGSAC Conf on Computer and Communications Security. New York: Assoc Computing Machinery, 2016: 270-282

[86] Zhang Lei, Hyperledger Avalon: Building the next wave of confidential applications [EB/OL]. Medium, 2019 [2021-05-25]. https://medium.com/iex-ec/hyperledger-avalon-building-the-next-wave-of-confidential-applications-54ba49dcd7e7

[87] Reed M G, Syverson P F, Goldschlag D M. Anonymous connections and onion routing [J]. IEEE Journal on Selected Areas in Communications, 1998, 16(4): 482-494

[88] Dingledine R, Mathewson N, Syverson P. Tor: The second-generation onion router [R]. Washington, DC: Naval Research Lab, 2004

[89] Freedman M J. Design and analysis of an anonymous communication channel for the free haven project [J/OL]. 2000 [2021-08-15]. https://citeseerx. ist. psu. edu/viewdoc/download?doi=10.1.1.26.8778

[90] Luo Junzhou, Yang Ming, Ling Zhen, et al. Anonymous communication and darknet: A survey [J]. Journal of Computer Research and Development, 2019, 56(1): 103-130 (in Chinese)
(罗军舟, 杨明, 凌振, 等. 匿名通信与暗网研究综述[J]. 计算机研究与发展, 2019, 56(1): 103-130)

[91] Poon J, Dryja T. The bitcoin lightning network: Scalable off-chain instant payments [J/OL]. 2016[2021-08-15]. https://www.https://www.bitcoinlightning.com/wp-content/uploads/2018/03/lightning-network-paper.pdf

[92] IBM. hyperledger-fabricdocs master [EB/OL]. [2021-08-15]. https://hyperledger-fabric.readthedocs.io/zh_CN/latest/test_network.html

[93] Li Qiang, Shu Zhanxiang, Yu Xiang, et al. Research on authentication Mechanism of blockchain System [J]. Journal of Command and Control, 2019, 5(1): 1-17 (in Chinese)
(李强, 舒展翔, 余祥, 等. 区块链系统的认证机制研究[J]. 指挥与控制学报, 2019, 5(1): 1-17)

[94] You Jiali, Xue Jiao, Wang Jinlin. A behavior cluster based availability prediction approach for nodes in distribution networks [C] //Proc of the 2013 IEEE Int Conf on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE, 2013: 2810-2814

[95] Tschorsch F, Scheuermann B. Bitcoin and beyond: A technical survey on decentralized digital currencies [J]. IEEE Communications Surveys & Tutorials, 2016, 18(3): 2084-2123

[96] Huang Butian, Liu Zhenguang, Chen Jianhai, et al. Behavior pattern clustering in blockchain networks [J]. Multimedia Tools and Applications, 2017, 76(19): 20099-20110

[97] Wang Peng, Wang Lin, Zhu Yuefei. Research and prevent sybil attackt in p2p networks [J]. Microelectronics & Computer, 2006, 23(4): 162-165 (in Chinese)
(王鹏, 王琳, 祝跃飞. 在 P2P 网络下 Sybil 攻击的研究与防范[J]. 微电子学与计算机, 2006, 23(4): 162-165)

[98] Yu Haifeng, Kaminsky M, Gibbons P B, et al. Sybilguard: Defending against sybil attacks via social networks [C] //Proc of the 2006 Conf on Applications, Technologies, Architectures, and Protocols for Computer Communications. New York: Association for Computing Machinery, 2006: 267-278

[99] Fang Weidong, Zhang Wuxiong, Yang Yang, et al. A resilient trust management scheme for defending against reputation time-varying attacks based on BETA distribution [J]. Science China Information Sciences, 2017, 60(4): No. 040305

[100] Qian Hong, Yu Yang. On sampling-and-classification optimization in discrete domains [C] //Proc of the 2016 IEEE Congress on Evolutionary Computation (CEC). Piscataway, NJ: IEEE, 2016: 4374-4381

[101] Liu Mingda, Shi Yijuan. Remote attestation model based on blockchain [J]. Computer Science, 2018, 45(2): 48-52 (in Chinese)
(刘明达, 拾以娟. 基于区块链的远程证明模型[J]. 计算机科学, 2018, 45(2): 48-52)

[102] Yuan Yong, Wang Feiyue. Editable blockchain: Models, techniques and methods [J]. Acta Automatica Sinica, 2020, 46(5): 831-846 (in Chinese)
(袁勇, 王飞跃. 可编辑区块链: 模型, 技术与方法[J]. 自动化学报, 2020, 46(5): 831-846)

[103] Mühle A, Grüner A, Gayvoronskaya T, et al. A survey on essential components of a self-sovereign identity [J/OL]. Computer Science Review, 2018, 30: 80-86 [2020-08-15]. https://doi.org/10.1016/j.cosrev.2018.10.002

[104] Tobin A, Reed D. The inevitable rise of self-sovereign identity [J]. The Sovrin Foundation, 2016, 29 (2016) [2021-08-15]. https://sovrin.org/wp-content/uploads/2017/06/The-Inevitable-Rise-of-Self-Sovereign-Identity.pdf

[105] Bernabe J B, Canovas J L, Hernandez-Ramos J L, et al. Privacy-preserving solutions for blockchain: Review and challenges [J/OL]. IEEE Access, 2019, 7: 164908-164940 [2020-08-15]. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8888155

[106] Batubara F R, Ubacht J, Janssen M. Challenges of blockchain technology adoption for e-government: A systematic literature review [C] //Proc of the 19th Annual Int Conf on Digital Government Research: Governance in the Data Age. New York: ACM, 2018: 1-9

[107] Sharma P K, Moon S Y, Park J H. Block-VN: A distributed blockchain based vehicular network architecture in smart City [J]. Journal of Information Processing Systems, 2017, 13(1): 184-195

[108] Sporny M, Longley D. Verifiable claims data model and representations [EB/OL]. [2020-08-15]. https://www.w3.org/TR/vc-data-model/

[109] W3C. Decentralized Identifiers (DIDs) v1.0 [EB/OL]. [2021-04-22]. https://w3c.github.io/did-core/

[110] Uport. Uport [EB/OL]. [2021-08-15]. https://www.uport.me/

[111] Windley P, Reed D. Sovrin: A protocol and token for self-sovereign identity and decentralized trust [J]. Whitepaper, The Sovrin Foundation, 2018[2021-08-15]. https://sovrin.org/library/sovrin-protocol-and-token-white-paper/

[112] EUROPA. TOC-EN-EUR-Lex [EB/OL]. [2021-04-22]. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

[113] Office of the Central Cyberspace Affairs Commission. Provisions on the management of blockchain information services [EB/OL]. [2021-08-15]. http://www.cac.gov.cn/2019-01/10/c_1123971164.htm (in Chinese)
(中共中央网络安全和信息化委员会办公室.区块链信息服务管理规定[EB/OL]. [2021-08-15]. http://www.cac.gov.cn/2019-01/10/c_1123971164.htm)

Wang Chenxu, born in 1986, PhD, associate professor, PhD supervisor. His main research interests include cross blockchain technology, data privacy-preserving, data mining, and network security.
王晨旭,1986 年生,博士,副教授,博士生导师.主要研究方向为区块链跨链技术、数据隐私保护、数据挖掘与网络安全.

Cheng Jiacheng, born in 1996. Master candidate. His main research interest is data privacy-preserving of blockchain.
程加成,1996 年生,硕士研究生,主要研究方向为区块链数据隐私保护.

Sang Xinxin, born in 1998. Master candidate. His main research interest is data privacy-preserving of blockchain.
桑新欣,1998 年生,硕士研究生,主要研究方向为区块链数据隐私保护.

Li Guodong, born in 1974. Research Fellow. His main research interests include network security and management.
李国栋,1974 年生,研究员.主要研究方向为网络安全与管理.

Guan Xiaohong, born in 1955. Professor and PhD supervisor. His main research interests include network information security, networked systems, optimal scheduling of power systems, resource bidding and game theory analysis.
管晓宏,1955 年生,教授,博士生导师.主要研究方向为网络信息安全、网络化系统、电力系统优化调度、资源竞标和博弈理论分析.