区块链数据隐私保护综述

闫朝文

计算机 2001 2120201848

摘　要：区块链技术作为分布式总账技术、智能合约基础平台、分布式新型计算范式，应用场景十分广泛，其独特优势能解决诸多行业痛点，并驱动新一轮技术变革和应用变革。但是区块链技术在实际应用中面临着严重的数据隐私泄露问题，极大限制了区块链的发展，如何在区块链平台上对交易数据进行隐私保护已成为研究者关注的重点问题之一。本文基于区块链数据隐私泄露的角度，详细分析了区块链各技术要点面临的隐私泄露问题，探索了当前区块链数据隐私保护的解决方案，并分析了联邦学习、安全多方计算这两种隐私保护模式的新型方案，结合目前区块链数据隐私脑虎研究的最新进展，为区块链安全领域相关人员未来的研究工作提供了一定的参考，并展望了未来区块链数据隐私保护的研究方向。

关键字：区块链；隐私保护；比特币；联邦学习；多方安全计算；

1 区块链数据隐私保护概述

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。随着区块链技术不断演进，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。然而随着个人隐私数据保护意识的觉醒，大众及组织机构对现有区块链数据的弱隐私性愈感不安[1]。

区块链是具有普适性的底层技术框架，其需要能够满足非常复杂的商业逻辑及企业间合作时的数据隐私保护。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，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[2]。

以比特币为例，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，但这也意味着恶意节点能够获取所有账本息。Neudecker 等通过对比聚类的区块链信息与泛洪比特币网络过程中提取的 IP 地址，得出了小部分聚类的区块链信息地址与 IP 地址有明显的关联，可以借此 IP 地址找出用户敏感信息。Goldfeder[3]等人的研究表明，比特币交易中交易接收方可以轻松将比特币的支付流通信息与用户 Cookies 相关联，从而去除比特币交易的匿名性，迫使其暴露出用户的真实身份。

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和源于区块链自身体系结构的隐私泄露问题正在逐步暴露。本文旨在对区块链技术发展现状、数据隐私威胁、数据隐私保护等攻击类型和保护方法进行综述，希望能给目前区块链数据隐私保护的相关研究提供一定的借鉴。

2 区块链数据隐私威胁

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、地址、身份证号等隐私信息

2013年，Androulaki等人[4]根据区块链钱包的应用特征提出了挖局找零地址的方法：如果一个交易拥有两个输出，其中一个为已出现过的地址，另一个为新地址，则将新地址视为找零地址，并提出了假设2，即交易的找零地址和输入地址为同一用户持有。为了检验假设的正确性，Androulaki等人在大学中构建了模拟的密码货币使用环境，通过搜集用户使用记录，利用假设1和假设2进行分析，实验结果表明接近40%的用户数据信息可以被揭露出来。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.

此类攻击方式利用了比特币系统自身漏洞，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，Meiklejohn等人[5]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，然后分类用户，从而识别出同一个用户的不同地址。

2.2 网络隐私及威胁

网络层关系整个底层的通信过程，在区块链中起着举足轻重的作用。区块链系统通过去中心化的P2P网络进行节点间通信，而在非许可链系统中的网络不存在准入限制，这在增强了扩展性的同时也带来了潜在的风险。攻击者可以任意部署节点，监听网络中各节点隐私信息以及网络通信信息，甚至尝试对正常节点发起攻击[6]。

区块链网络隐私主要包含一下两种：

其一是节点隐私，节点自身的隐私内容，包含节点网络 IP、软件版本、服务器系统等隐私信息；

其二是通信隐私，节点间通信隐私内容,包含节点间通信的数据内容以及通信流量情况。

Reid[7] 等人尝试利用 Bitcoin Faucet 公开的区块链地址及 IP 地址对应关系进行分析，揭露比特币用户与实际物理位置的对应关系，该尝试只涉及到很少的节点。Bitnodes 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数据隐私安全。IP 地址分布信息如图 1 所示，截止日期为 2019 年 8 月 1 日。

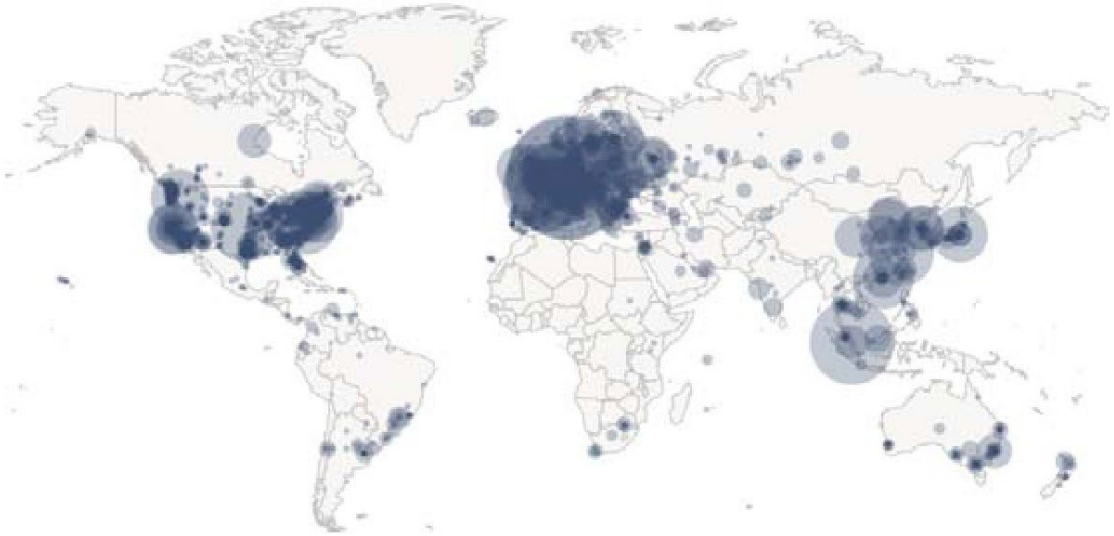
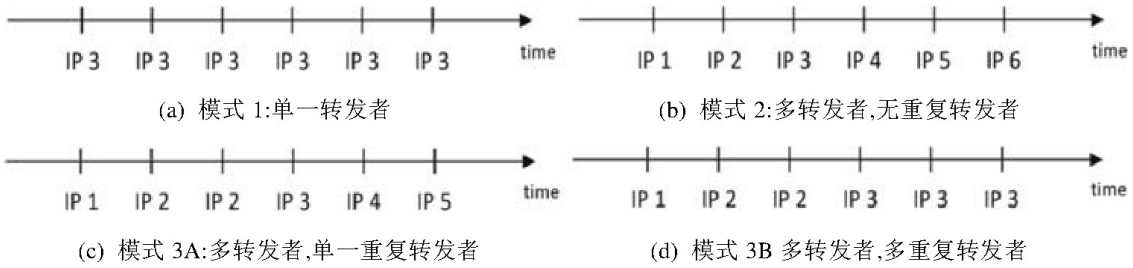


图 1 世界范围内比特币节点分布

尽管区块链网络采用洪水广播的方式保护实际发起人，但是攻击者通过大量探测可以将区块链中广播的数据与实际节点关联起来，有很大概率找出消息的真实发起节点，Kaminsky[8]在黑帽大会上的假设第一次接收到消息时的来源节点即为该消息的真实发起节点。

Koshy 等人在 Kaminsky 提出假设的基础上进行了完善，提出了区块链网络中消息传播的 4 种模式(如图 2 所示)以及对应的真实发起者假设。这一攻击方式通过监听消息传播模式，分析真实发起节点，将 IP 地址与消息中包含的链上地址对应，威胁通信隐私与用户身份隐私。



(a) 模式 1:单一转发者　(b) 模式 2:多转发者,无重复转发者

(c) 模式 3A:多转发者,单一重复转发者　(d) 模式 3B 多转发者,多重复转发者

图 2 4 种消息传播模式

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。

2.3 跨数据隐私及威胁

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，导致跨链数据隐私保护面临进一步的挑战[9]。

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[10]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。

3 区块链数据隐私保护

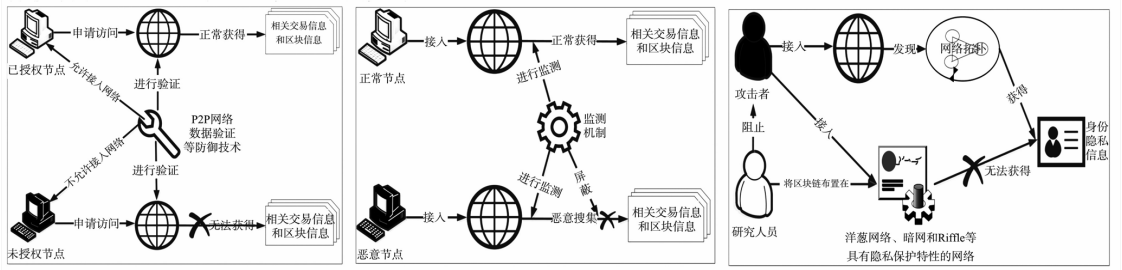
3.1 网络层数据隐私保护

网路层数据隐私保护技术大致分为 3 类，具体如图 3 所示。

1）限制接入 对区块链的节点进行管理授权点，未经授权的节点不能访问网络并获取相关的交易信息和阻止信息，例如 P2P 网络技术、数据验证机制等。

2）恶意节点监测和屏蔽 在公共链架构中无法直接约束节点对网络进行访问，但是检测机制可以被用来发现恶意节点并将其加入禁止访问名单，可以防止恶意节点继续获取隐私数据。

3）网络层数据混淆 为了预防攻击者利用网络拓扑来获取身份隐私信息，有些研究者提出在拥有隐私保护功能的网路上放置区块链，例如洋葱网络[12]、暗网和 Riffle[13]。



(a) 限制接入　(b) 恶意节点监测和屏蔽　(c) 网络层数据混淆

图 3 网络层数据隐私主要保护技术

洋葱路由是一种保护网络通信的真实发送和接受节点，达成匿名通信的网络链路协议。洋葱路由通过对传输消息进行多层加密，使得路由中间节点只能知道前继节点和后继节点的 IP 地址，不能获取消息的真实发送节点 IP 与接收节点 IP，保护消息发送方和接收方的真实 IP 不被攻击者知晓。

洋葱路由的代表项目为 Tor 网络[14]，Tor 网络建立发起者到接受者的链路主要包含三个阶段：节点选择阶段、链路建立阶段、消息传输阶段，如图 4 所示。

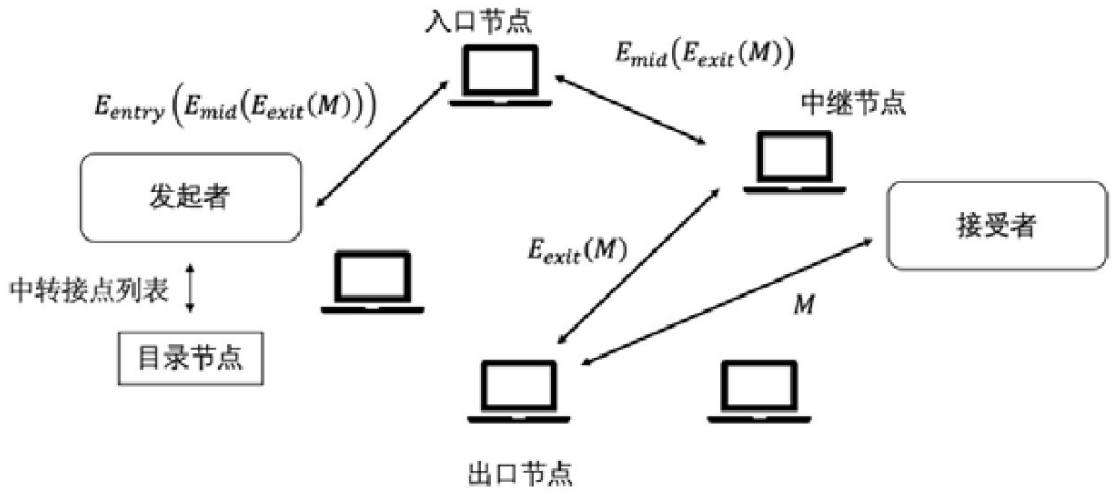


图 4 Tor 网络

为了增强用户隐私保护，部分比特币钱包内置了 Tor 网络配置，使得运行比特币钱包的节点可以通过 Tor 节点与比特币网络连接，从而保护节点的 IP 等隐私信息。

3.2 应用层数据隐私保护

当用户使用区块链应用程序时，其交易数据和身份隐私就可能泄露。攻击者主要是利用用户不规范的操作和区块链服务商的漏洞搜集交易隐私和身份隐私。因此应用层隐私保护的核心是提高用户隐私安全意识和提升相关区块链应用程序的隐私防护质量。

用户可以采用的防御方法通常有两种：

1）使用具有隐私保护机制的区块链应用。以比特币为例，比特币在隐私保护方面存在明显缺陷，攻击者可以通过多种方法获得身份隐私和数据隐私。在这种背景下，出现了许多隐私保护效果更好的替代货币，例如达世币、门罗币、零币 。

达世币-Dash[15]，基于 POW 共识机制，它的两种节点“Masterodes（主节点）”和”Miners（矿工节点）”均在网络上使用。达世币的匿名性是通过混币技术所实现， 指的是将几笔交易的全部代币混在一起，然后再发往不同的收款方。 这样每一次混币，都会使得追查到交易发起方地址的难度呈指数型增加。除此之外，达世币还可以做到防篡改的即时交易。它利用主网中随机六个主节点暂扣支付 DASH 的方式，先确定完成支付过程，使达世币即时到账，然后才在网络中进行交易广播， 等待六次确认之后，收款方才能够继续使用刚收到的达世币。

门罗币-Monero[16]，采用环签名技术模糊交易的输入地址来增加攻击者分析资金来源的难度。由于交易中牵涉的地址和金额(含发送方和接收方)在账本中是独有的，环签名过程也不需要其他节点的参与，避免了同类混币服务面临的拒绝服务攻击和中间节点泄露混币过程等威胁。

零币-Zcash[17]，是目前隐私保护效果最好的数字货币。零钞利用零知识证明(zero knowledge proof)让用户只是通过和加密货币本身进行交互来隐藏交易信息。用户还可以用赎回操作将混币池中的零钞提取出来，赎回操作是把一个承诺重新换为零钞。矿工也并不知道哪个承诺被赎回换为零钞。因此甚至可以说不用向任何人转账，仅仅是把一个零钞放到混币池中再赎回，它的来源都不可追踪。

2） 使用具有隐私保护机制的区块链程序。不同的区块链程序在隐私保护方面具有不同的特点，需要采用针对性的保护方法。常见的区块链隐私保护程序有冷钱包、多重签名技术等。

多重签名，徐朝东等人[18]基于罚金机制的公平交换技术和区块链技术，提出了基于区块链的有序多重签名方案。方案将区块链技术与有序多重签名方案相结合，保证了签名过程完全透明，而且计算效率得到显著提高。

4 区块链隐私保护的新兴方案

4.1 异步联邦学习

联邦学习（federated leaning）[19]是顺应大数据时代和人工智能技术发展而兴起的一种协调多个参与方共同训练模型的机制，它允许各个参与方将数据保留在本地，在打破数据孤岛的同时保证参与方对数据的控制权。

但是随着参与训练的角色增多、能力增强，联邦学习也由此面临着与集中式机器学习不同的隐私泄露风险。联邦学习引入了大量参数交换过程，不仅和集中式训练一样受到模型使用者的威胁，还可能受到来自不可信的参与设备的攻击。

刘艺璇等人[20]基于联邦学习的隐私保护机制和技术，为其设计了多种保护措施，隐私保护对象明确为中心或本地，将联邦学习隐私保护算法分为 3 种主要类型：中心保护、本地保护、中心与本地同时保护策略。3 种保护策略的区别如图 5 所示。

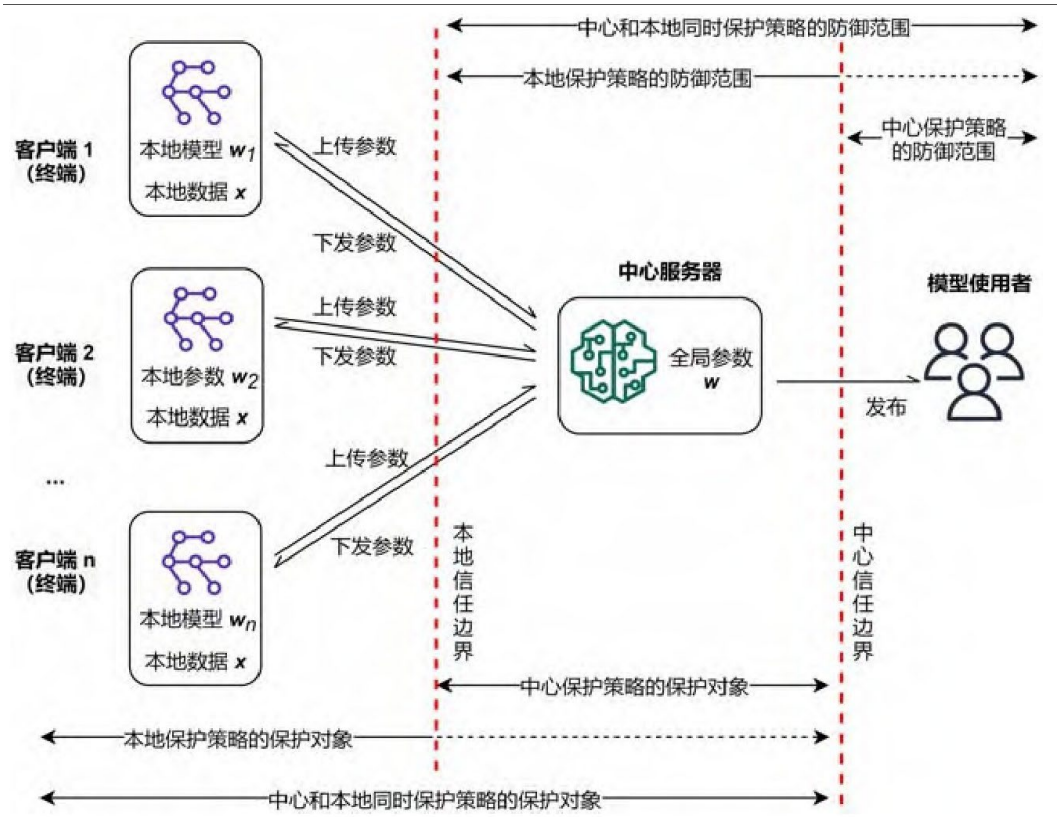


图 5 隐私保护策略的信任边界

中心与本地同时保护的策略能够为联邦学习提供全周期的保护，但是需要以中心及本地策略为基础；而中心和本地在满足一定的信任条件或可用性要求时，也有其适用场景。中心扰动

方法无法防御内部攻击，但是迁移方法、安全计算与扰动结合方法、安全混洗方法都需要以中心扰动为基础设计，中心扰动方法的研究价值依然很高。

4.2 安全多方计算

安全多方计算（MPC），是一种在参与方不共享各自数据且没有可信第三方的情况下安全地计算约定函数的技术和系统。通过安全的算法和协议，参与方将明文形式的数据加密后或转化后再提供给其他方，任一参与方都无法接触到其他方的明文形式的数据，从而保证各方数据的安全。

张硕等人[21]利用安全比较协议来构建多用户多关键词公钥可搜索加密算法，能够比较参与方隐私输入的大小，且不泄漏参与方的隐私输入，满足可搜索加密中需要判断云服务器端关键词与用户关键词是否匹配并且不泄漏关键词的需求；之后基于线性秘密共享协议设计广播加密，将包含主私钥部分信息的私密值部署到区块链网络，恢复该私密值需要区块链网络提供协助。从而完成广播加密的解密过程；最后利用旷工来执行安全计算协议，使用此过程中防寒生的零知识证明协议的证据保证安全多方计算的正确性和安全性，使原本无意义的哈希函数计算变为功能丰富的安全多方计算。

冯琦等人[22]针对三方协同计算协议扩展性差的问题，将多方计算模式与机器学习算法相结合，设计了多层次的，分布式的，具有数据隐私保护的深度神经网络训练方案。该方案通过引入混合训练架构，面向两种通信和计算环境封闭设计适应性的安全多方计算子协议，达到效能代价平衡。同时在准备阶段预计算部分关键数据，减少分布式训练过程中的计算量和通信量。

5 结论与展望

随着区块链技术的成熟以及各场景下区块链系统落地数量的逐步增长，其面临的安全问题将从多角度多维度体现出来。特别是在公共领域，用户的数据信息很容易被第三方监控、保存和利用，数据隐私保护问题毫无疑问将成为业内研究者的重点研究方向。

区块链势必将和云计算、人工智能、大数据等互联网前沿技术互相融合，那么将会带来更多的安全问题[23]。本文结合区块链技术的历史发展进程，介绍了几个容易产生隐私泄露问题的区块链层面，并详细介绍了几类隐私保护机制的原理、特征以及不同的实现方式。

现有的各类隐私保护机制及实现技术从不同方面保护区块链隐私，因而在实际考虑隐私保护的区块链系统中，通常综合多种技术以达到更全面的保护效果，区块链隐私保护仍然等待着更多广泛、深入的研究。

参考文献

[1]王晨旭,程加成,桑新欣,李国栋,管晓宏.区块链数据隐私保护：研究现状与展望[J].计算机研究与发展,2021,58(10):2099-2119

[2]祝烈煌,高峰,沈蒙,李艳东,郑宝昆,毛洪亮,吴震.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

[3]刘峰,杨杰,齐佳音.区块链密码学隐私保护技术综述[J].网络与信息安全学报,2022,8(04):29-44

[4] Androulaki E, Karame GO, Roeschlin M, Scherer T, Capkun S. Evaluating user privacy in Bitcoin. In:Proc. of the Int'l Conf. on Financial Cryptography and Data Security. 2013. 34-51

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

[6]张奥,白晓颖.区块链隐私保护研究与实践综述[J].软件学报,2020,31(05):1406-1434.DOI:10.13328/j.cnki.jos.005967

[7]Reid F, Harrigan M. An analysis of anonymity in the Bitcoin system. In:Proc. of the Security and Privacy in Social Networks.2013. 197-223

[8]Kaminsky D. Black Ops of TCP/IP 2011. 2011

[9]王丹. 基于区块链应用中用户身份和交易数据隐私保护研究[D].烟台大学,2022.DOI:10.27437/d.cnki.gytdu.2022.000376.

[10]Raze Network.Trustless Privacy on Polkadot[EB/OL].[2021-08-15]

[11]马彦兵.基于洋葱网络的流量确认攻击及其防御技术研究[D].北京：北京邮电大学, 2018.MA Yanbing.Research on traffic confirmation attack and defense based on tor network[D]. Beijing:Beijing University of Posts and Telecom,2018.

[12]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.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

[13]DALA-CORTE R B,FRIES L.Inter and intraspecific variation in fish body size constrains microhabitat use in a subtropical drainage[J].Environmental Biology of Fishes,2018,101(7):1205-1217

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

[15]米沃奇.被"抛弃"的比特币与前赴后继的暗网新宠[J].电脑知识与技术(经验技巧),2018(08):94-97

[16]林定康,颜嘉麒,巴·楠登,符朕皓,姜皓晨.门罗币匿名及追踪技术综述[J].计算机应用,2022,42(01):148-156.

[17]张宪,蒋钰钊,闫莺.区块链隐私技术综述[J].信息安全研究,2017,3(11):981-989.

[18]徐朝东. 基于区块链的多重签名研究[D].南京邮电大学, 2021.DOI: 10.27251/d.cnki.gnjdc.2021.000099.

[19]朱建明,张沁楠,高胜,丁庆洋,袁丽萍.基于区块链的隐私保护可信联邦学习模型[J].计算机学报,2021,44(12):2464-2484.

[20]刘艺璇,陈红,刘宇涵,李翠平.联邦学习中的隐私保护技术[J].软件学报,2022,33(03):1057-1092.DOI:10.13328/j.cnki.jos.006446.

[21]张硕. 安全多方计算协议及其应用研究[D].北京邮电大学,2021.DOI:10.26969/d.cnki .2021.000212.

[22]冯琦. 基于安全多方计算的数据隐私保护技术研究[D].武汉大学,2021.DOI:10.27379/d.cnki.gwhdu.2021.000902.

[23]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