计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2017.20170471
54(10):2170-2186,2017

区块链隐私保护研究综述

祝烈煌[1] 高 峰[1] 沈 蒙[1] 李艳东[1] 郑宝昆[1,2] 毛洪亮[3] 吴 震[3]
[1](北京理工大学计算机学院 北京 100081)
[2](中国政法大学 北京 102249)
[3](国家计算机网络应急技术处理协调中心 北京 100029)
(liehuangz@bit.edu.cn)

Survey on Privacy Preserving Techniques for Blockchain Technology

Zhu Liehuang[1], Gao Feng[1], Shen Meng[1], Li Yandong[1], Zheng Baokun[1,2], Mao Hongliang[3], and Wu Zhen[3]
[1](*School of Computer Science, Beijing Institute of Technology, Beijing* 100081)
[2](*China University of Political Science and Law, Beijing* 102249)
[3](*National Computer Network Emergency Response Technical Team/Coordination Center of China (CNCERT/CC), Beijing* 100029)

Abstract Core features of the blockchain technology are "de-centralization" and "de-trusting". As a distributed ledger technology, smart contract infrastructure platform and novel distributed computing paradigm, it can effectively build programmable currency, programmable finance and programmable society, which will have a far-reaching impact on the financial and other fields, and drive a new round of technological change and application change. While blockchain technology can improve efficiency, reduce costs and enhance data security, it is still in the face of serious privacy issues which have been widely concerned by researchers. The survey first analyzes the technical characteristics of the blockchain, defines the concept of identity privacy and transaction privacy, points out the advantages and disadvantages of blockchain technology in privacy protection and introduces the attack methods in existing researches, such as transaction tracing technology and account clustering technology. And then we introduce a variety of privacy mechanisms, including malicious nodes detection and restricting access technology for the network layer, transaction mixing technology, encryption technology and limited release technology for the transaction layer, and some defense mechanisms for blockchain applications layer. In the end, we discuss the limitations of the existing technologies and envision future directions on this topic. In addition, the regulatory approach to malicious use of blockchain technology is discussed.

Key words blockchain; privacy-preserving; peer-to-peer networking; clustering analysis; Bitcoin

摘 要 区块链技术的核心特征是"去中心化"和"去信任化",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,可以有效构建可编程货币、可编程金融和可编程社会,势必将对金融及其他领域带来深远影响,并驱动新一轮技术变革和应用变革.但是区块链技术在提高效率、降低成本、提高数据安

收稿日期:2017-06-11;**修回日期**:2017-08-03
基金项目:国家重点研发计划项目(2016YFB0800301);国家自然科学基金项目(61602039);北京市自然科学基金项目(4164098)
This work was supported by the National Key Research and Development Program of China (2016YFB0800301), the National Natural Science Foundation of China (61602039), and the Beijing Natural Science Foundation (4164098).
通信作者:沈蒙(shenmeng@bit.edu.cn)

全性的同时,也面临严重的隐私泄露问题,得到研究者的广泛关注.将介绍区块链技术架构,定义区块链技术中身份隐私和交易隐私的概念,分析区块链技术在隐私保护方面存在的优势和不足,并分类描述现有研究中针对区块链隐私的攻击方法,例如交易溯源技术和账户聚类技术;然后详细介绍针对区块链网络层、交易层和应用层的隐私保护机制,包括网络层恶意节点检测和限制接入技术、区块链交易层的混币技术、加密技术和限制发布技术,以及针对区块链应用的防御机制;最后,分析了现有区块链隐私保护技术存在的缺陷,展望了未来发展方向.此外,还讨论针对恶意使用区块链技术的监管方法.

关键词 区块链;隐私保护;对等网络;聚类分析;比特币

中图法分类号 TP391

区块链技术具有“去中心化”和“去信任化”等特点,能够不依赖第三方可信机构在陌生节点之间建立点对点的可信价值传递,有助于降低交易成本,提高交互效率,有非常广阔的应用前景,被认为是引领信息互联网向价值互联网转变的关键技术[1].

在数字货币领域,比特币和类似比特币的新型数字货币发展迅速.ARK 投资公司和 coinbase 公司 2017 年发布的联合报告指出:全球比特币用户数量超过 1 000 万,每天交易金额超过 1.5 亿美元[2].Gartner 的预测指出,到了 2022 年,以区块链为中心的相关交易将高达 100 亿美元[3].此外,各国央行逐渐支持数字货币,甚至计划发行法定数字货币.在金融行业,银行和金融机构希望利用区块链技术改造传统金融体系,降低数据维护成本.在能源行业,区块链的去中心化、开源共享以及智能管理等特性与未来能源变革有着天然的契合性,能够在能源互联网、智能电网、碳交易、电动汽车、智能交通等领域发挥重要作用[4].在文化行业,利用区块链技术数据不可更改、公信力高的特性,可以开展证书存储、数字产权保护、文物鉴定等众多业务.

随着区块链技术不断发展和广泛应用,其面临的隐私泄露问题越来越突出,必须得到充分重视.相对于传统的中心化架构,区块链机制不依赖特定中心节点处理和存储数据,因此能够避免集中式服务器单点崩溃和数据泄露的风险.但是为了在分散的区块链节点中达成共识,区块链中所有的交易记录必须公开给所有节点,这将显著增加隐私泄露的风险.例如,在数字货币应用中,分析人员通过分析交易记录可以获得用户的交易规律,甚至能够推测出用户的身份信息和位置信息[5].在金融应用中,如果分析人员获得全部的交易记录,既可以追溯个体账户的交易细节,也可以分析宏观的金融趋势,这些信息既属于用户的隐私信息,也属于公司的核心数据.在能源行业中,区块链技术通常被用于实现点对点的能源交换[6],这种情况下区块链交易数据有可能泄露能源传输等敏感信息,对个人安全和国家安全造成威胁.因此,区块链技术在走向实用之前,必须解决隐私泄露问题.

然而,区块链技术与传统 IT 架构存在显著区别,很多传统的隐私保护方案在区块链应用中不适用.传统 IT 架构中,数据通常存储在中心化服务器,隐私保护的重点是确保数据不被外泄.因此,管理者可以通过提高中心节点的防御能力来抵抗各种攻击,例如使用高性能服务器、部署入侵检测设备、安装专用的数据防泄密软件等.区块链技术中,数据存储在分散的节点,没有统一的管理者,节点的性能和安全能力参差不齐,攻击者很容易攻陷其中一些节点.此外,攻击者甚至可以伪装成合法节点直接获得交易数据.因此,区块链中隐私保护的重点是确保交易的匿名性,即攻击者无法通过分析交易数据获得用户的身份信息,这种安全需求需要针对性的隐私保护机制.区块链技术中采用了特殊的信息传递机制、共识机制和激励机制,这也给隐私保护带来了新的机遇和挑战.

区块链技术面临隐私泄露风险,传统的隐私保护技术又不适用,因此分析区块链隐私泄露缺陷、研究针对性的隐私保护方法具有重要意义.目前已经出现了很多针对区块链隐私的防护方法,它们能够从网络层、交易层等不同的角度防范窃取隐私的攻击行为.针对目前发展现状,本文对现有典型的攻击方法和防御机制进行回顾与总结,希望能给当前及未来的相关研究提供一定的参考与帮助.

1 区块链隐私保护背景知识

1.1 区块链技术概述

区块链是从比特币底层技术衍生出来的新型技术体系,最早的定义来自于中本聪在 2009 年发表的论文[7],之后区块链的内涵和外延发生了很多改变,

目前仍然在不断演变. 区块链技术在架构上通常被分为数据层、网络层、共识层、激励层、合约层和应用层. 但是随着区块链技术的快速发展,区块链的架构也在不断变化. 很多传统的模块被弱化,甚至被取消,例如在联盟链和私有链技术中已经不需要激励层. 因此,通过分析区块链技术的本质特征和目前的发展趋势,我们将区块链技术的架构分为 3 个层次,如图 1 所示:

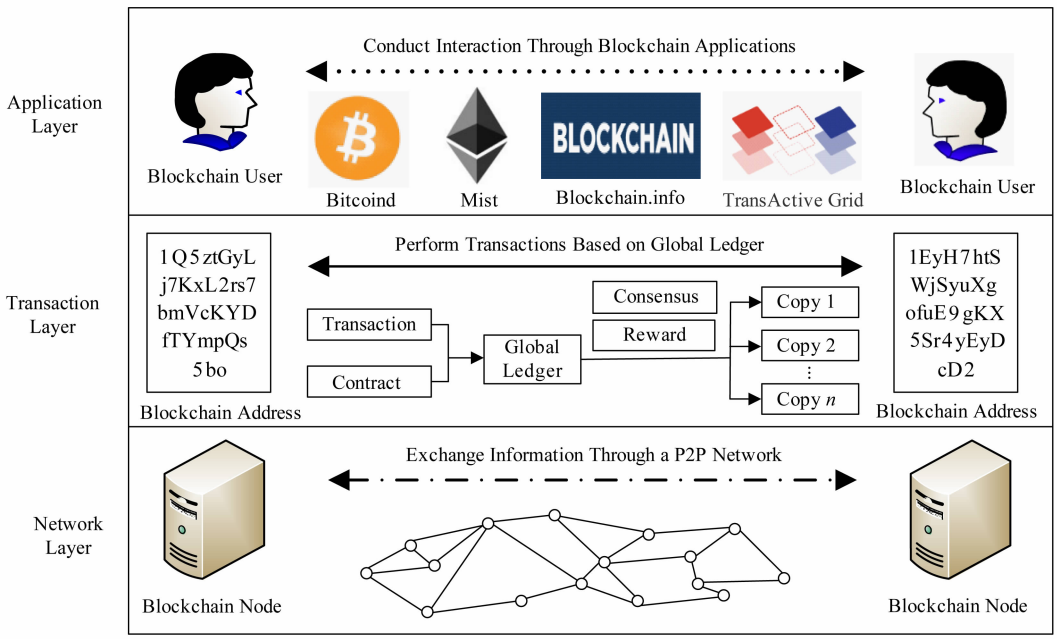


Fig. 1 The architecture of blockchain technology

图 1 区块链技术架构

1) 网络层. 网络层的核心任务是确保区块链节点之间可以通过 P2P 网络进行有效通信. 主要内容包括区块链网络的组网方式和节点之间的通信机制.

区块链网络采用 P2P 组网技术,具有去中心、动态变化的特点. 网络中的节点是地理位置分散但是关系平等的服务器,不存在中心节点,任何节点可以自由加入或者退出网络. 目前规模最大的区块链网络是比特币网络. 比特币网络建立在公共互联网之上,节点来自全球各地,每天对外提供服务的节点数量平均为 5 400 个[8],总体节点数量(包括不对外提供服务的节点)估计为 10 万个左右.

区块链节点之间的通信类型主要分为 2 种:

① 为了维持节点与区块链网络之间的连接而进行的通信,通常包括索取其他节点的地址信息和广播自己的地址信息(地址信息是指 TCP/IP 中的 IP 地址和端口号). 节点新加入区块链网络时,首先读取硬编码在客户端程序中的种子地址并向这些种子节点索取其邻居节点地址,然后通过这些地址继续搜索更多的地址信息并建立连接,直到节点的邻居节点的数量达到稳定值. 此后,节点会定期通过 ping 等方式验证邻居节点的可达性,并使用新的节点替代不可达节点. 此外,为了保证新节点的信息被更多节点接收,节点将定期向自己的邻居节点广播自己的地址信息.

② 为了完成上层业务而进行的通信,通常包括转发交易信息和同步区块信息(交易和区块是区块链中的数据结构,将在交易层介绍). 节点转发交易信息时采用中继转发的模式. 始发节点首先将交易转发给邻居节点,邻居节点收到交易后再转发给自己的邻居节点,以此类推,逐渐传遍整个网络. 同步区块信息采用请求响应的模式. 节点首先向邻居节点发送自己的区块高度(类似于 ID),如果小于邻居节点的高度则索取自己欠缺的区块,如果大于邻居节点的高度则邻居节点将反向索取区块信息. 所有节点都不断地和邻居节点交换区块信息,从而保证整个网络中所有节点的区块信息保持同步.

2) 交易层. 交易层实现区块链的核心业务,即在 2 个“地址”之间进行可靠的、具有公信力的数据传递. 主要内容包括:地址格式、交易格式、智能合约、全局账本、共识机制和激励机制.

区块链中的“地址”,类似于银行卡账号,是用户参与区块链业务时使用的假名. 通常是在用户的控制下利用公钥加密算法(例如 ECC)生成. 其中生成的公钥信息将用于交易的输入地址或者输出地址,

私钥信息由用户自己保存,用于对交易签名.2 种常见的区块链地址如下所示:

① 比特币地址:

“1DAY1DUpbBdGLkkFYj32J5g4h9X2zsxDv5”

② 以太坊地址:

“02B51B20185c04D1CbDA2996dFA02AF2D308EeEa”

区块链中的“交易”记录了用户之间数据交互的过程.通常包括输入地址、输出地址、交易内容等信息.交易内容在数字货币应用中主要包括交易的金额,在其他应用中可能代表一个字符串或者一个证书 ID.例如在基于区块链技术的数据存储应用 STORJ[9] 中,交易的内容主要包括存储数据的 Hash 值.我们以使用最广泛的区块链应用比特币为例,介绍交易层的数据格式.

图 2(a)展示一个简化的比特币交易格式,图 2(b)展示 2 个交易例子.从交易格式部分可以看出,每个比特币交易都有一个交易 Hash(txhash).此 Hash 值是针对整个交易内容计算得到,唯一指向此条交易.因此,在比特币中交易 Hash 通常作为交易的 ID.交易的正文主要包括 2 部分:输入地址信息和输出地址信息.输入地址信息记录此次交易发送方的账户信息,值得注意的是输入地址信息中并不是记录发送者的账号,而是记录输入资金的来源(pre-txhash),即通过来源交易 Hash 指定全局账本中的一条交易,通过索引信息(index)指定交易中对应的输出地址,并通过签名信息(sign)证明用户对这笔资金的所有权.输出地址信息中记录此次交易接收方的账户信息,包括输出地址(account)和输出金额(amount).输出地址是由用户自行生成的公钥信息经过字符变换得到的一串字符串.输出地址经过反向变换后可以得到公钥的 Hash,用于验证签名.交易实例部分介绍了 2 条交易.其中,编号为“10002”的交易中第“0”个输入地址中的来源交易 Hash 是“3306ff5a64d900937ad1429466fd2c8f”,同时索引为“1”.因此,可以确定此输入地址的真实账户是编号“10001”交易中第“1”个输出地址“1A1RmbbVoL4pnMZf”.通过这种设计,比特币中每一个交易的来源都可以验证真实性和合法性,攻击者无法伪造交易,也不能冒用其他人的账号进行交易.此外,交易中的输入账户和输出地址都是由用户自行创建,与身份信息无关,因此外部观察者不能直接根据交易记录推测交易者的身份信息.

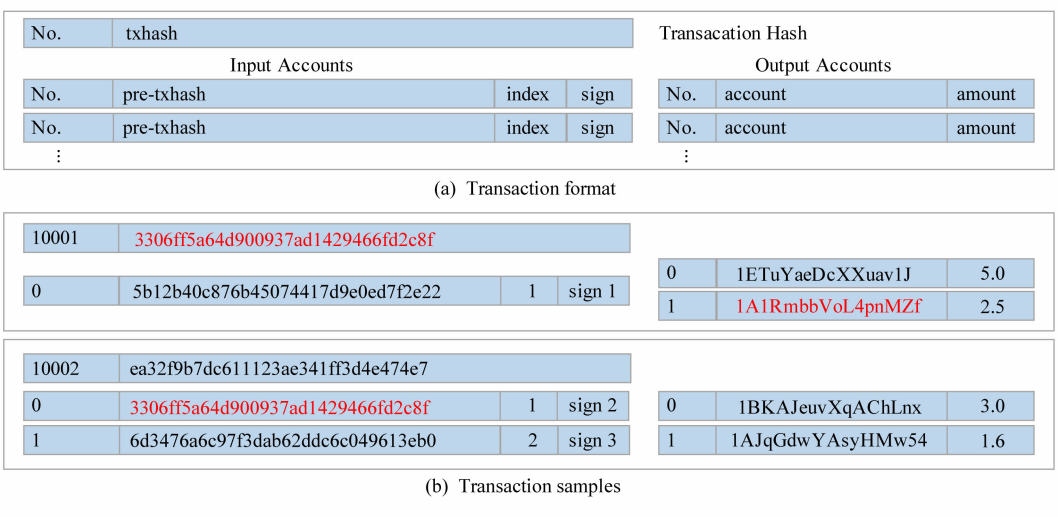


Fig. 2 The format of bitcoin transaction

图 2 比特币交易格式

合约是区块链中一种特殊的数据交换方式,能够提供比交易更灵活的数据交换.区块链中的合约的概念在不断变化.在比特币中,合约是交易中的脚本,用于实现延时交易、条件交易等复杂业务.在以太坊中,合约被扩展为图灵完备的脚本语言,可以实现更加复杂的功能.在超级账本(Hyperledger)中,合约直接用于实现区块链业务.同交易类似,合约的内容也将写入区块链全局账本.

全局账本是区块链中的数据存储结构,用于存储所有的交易记录、合约以及相关的参数信息.全局账本通常由“区块”构成,每个区块存储一定数量的交易信息以及针对这些交易的 Hash 值、时间戳等参数.区块之间按照时间关系通过区块 Hash 连接.全局账本实际上从初始区块到最新区块的数据链

条,这也是区块链名字的由来.全局账本由所有参与节点共同维护,每一个节点各自维护本地的全局账本,节点通过定期和邻居节点交换信息使全局账本保持同步.持有不同全局账本的节点不能参与后续的交易.

区块链技术采用共识机制保证所有合法节点维持的全局账本是相同的.常见的共识机制包括POW机制[10]、POS机制[11]、PBFT机制[12]等.POW机制的核心思路是设置一个数学难题,让参与节点求解难题,在求解过程中付出最大工作量(算力)的节点将被选择为记账节点,即由此节点生成新的区块.其他节点通过接受此区块更新自己的全局账本.通过选择一个特定用户记账,解决了多用户记账带来的数据不同步问题.但是此类共识机制将浪费大量的算力,同时导致记账权逐步被拥有大额算力的用户(例如矿池)垄断,带来很多的安全问题.POS机制通过使用币天(节点持有的数字货币和持有的天数)来选择记账节点,不需要消耗大量的计算资源,目前被很多应用作为POW的替代机制.此外,还有许多共识机制从速度、安全性等角度做出了改进,满足区块链技术在不同业务场景的需要.

为了鼓励更多用户参与共识,提高系统的安全性,最初的区块链技术中设置了激励机制奖励参与共识的用户.例如在比特币中,参与记账的节点被称为矿工,成功获得记账权的矿工节点将收到比特币作为奖励.但是随着区块链技术的发展,区块链的应用场景从公有链衍生到联盟链和私有链,在这些场景中节点是可控的,因此不需要设置额外的激励机制.

3) 应用层.应用层提供针对各种应用场景的程序和接口,用户通过部署在应用层的各种应用程序进行交互,而不用考虑区块链底层技术细节.目前典型的区块链应用包括数字货币应用、数据存证应用、能源应用等.

数字货币应用是区块链中最早出现的应用,除了比特币以外,目前出现了大量的竞争币,例如以太币、Zcash、门罗币等.区块链中的数字货币是由被称为矿工的节点在参与区块链共识机制时创建.普通用户可以通过交易所或者私下交易的方式购买数字货币.用户持有的数字货币可以在数字货币系统中进行交易,用于购买商品或者服务.目前已有很多大型的企业和商户支持数字货币交易,例如微软和亚马逊.

数据存证应用是区块链技术重要的应用方向.区块链中的全局账本具有不可篡改、抗攻击能力强的特点,非常适合存储重要的数据资料,例如知识产权文件、金融交易记录等.此类应用一般由区块链公司开发便捷的存证系统,普通用户通过存证系统上传文件资料,然后由系统完成后续写入区块链的流程.存证成功后,用户使用系统返回的交易Hash值查看存储记录.

1.2 区块链隐私定义

信息系统中隐私通常是指数据拥者不愿意被披露的敏感数据或者数据所表征的特性[13].

区块链技术为了在分散节点之间维持数据同步并对交易达成共识,必须公开一些信息,例如交易内容.另一方面,为了保护用户隐私,必须对一些敏感数据进行处理,减少隐私泄露的风险.通过分析区块链技术的特点,本文将区块链中的隐私分为2类.

定义1. 身份隐私.身份隐私是指用户身份信息和区块链地址之间的关联关系.

区块链地址是用户在区块链系统中使用的假名,通常作为交易的输入账号或者输出账号.地址的格式已在1.1节介绍.区块链系统中地址是由用户自行生成,与用户身份信息无关,用户创建和使用地址不需要第三方参与.因此,相对于传统的账号(例如银行卡号),区块链地址具有较好的匿名性.但是,用户在使用区块链地址参与区块链业务时,有可能泄露一些敏感信息,例如区块链交易在网络层的传播轨迹,这些信息有可能被用于推测区块链地址对应的真实身份.

定义2. 交易隐私.交易隐私是指区块链中存储的交易记录和交易记录背后的知识.

在早期的区块链数字货币应用中,交易记录通常是公开的,不需要额外的保护措施.但是随着区块链技术被应用到银行等金融领域,交易记录属于重要的敏感数据,需要采取额外措施限制非授权用户的使用.此外,交易记录通常能够反映一些敏感知识,有可能泄露用户的隐私.例如用户购物的交易记录能够反映用户的消费水平、生活状态等.

身份隐私和交易隐私是用户在使用区块链技术时需要重点保护的内容,这些信息一旦泄露有可能对用户造成危害.而且由于存储在区块链全局账本中的数据无法删除和篡改,即使用户发现部分地址或者交易数据已经曝光,也不能采取挽救措施.例如,传统领域中可以通过删除已曝光数据来减少隐私传播范围,但是在区块链中很难实行类似方案.因此,区块链系统应该更加重视隐私问题,提高隐私防护能力.

2 区块链隐私威胁

2.1 区块链隐私保护存在的优势和缺陷

区块链技术在隐私保护方面具有一些突出优势,能够解决一些中心化服务器面临的隐私泄露问题.因此已经被应用到许多需要保护隐私的场景,例如基于区块链技术的匿名投票.另一方面,区块链技术采用的去中心化架构和数据存储机制给隐私保护带来一些不利因素.表1分析了区块链技术在隐私保护方面存在的优缺点.

区块链技术在隐私保护方面的优势包括:

1) P2P网络很难实现网络窃听.区块链网络是一种P2P网络,节点之间采用中继转发的模式进行通信,传统网络中通过窃听网络流量发现用户之间通信关系的方法不适用.例如区块链网络中,当节点之间需要进行交易时,发送方首先将交易信息发给自己的邻居节点,收到信息的邻居节点再将信息转发给自己的邻居节点,以此类推,信息将逐渐广播到整个网络.接收方节点最终将从网络中收到交易信息,而不需要和发送方直接通信.因此,攻击者很难通过窃听发现网络中传播信息的真实来源和去向.

2) 区块链技术支持匿名交易.区块链交易中使用的地址(类似于银行卡账户)通常由用户自行创建和保存,不需要第三方参与,地址本身和用户身份信息无关.此外,区块链地址通常具有非常大的地址空间,出现碰撞的概率非常低,这使得用户可以为每次交易生成不同的地址,增强交易的匿名性.例如比特币地址对应的私钥空间是2^{256},用十进制表示是10^{77},而可见宇宙被估计只含有10^{80}个原子[14],因此比特币系统有充足的地址空间支持一次性地址策略.

3) 去中心化架构能够有效应对网络攻击.采用区块链技术的应用程序通常是去中心化架构,不需要在中心服务器上存储账户、密码等敏感信息,能够避免传统服务器被攻击而导致的数据泄露风险.

区块链技术在隐私保护方面的不足之处包括:

1) 区块链网络中的节点容易遭受攻击.区块链网络中的节点通常是个人电脑,和传统网络架构中的专用服务器相比性能低、抗攻击能力差.此外,在中心化架构中,管理者只需针对一台或者少数几台服务器进行重点保护.而在区块链网络中,所有节点地位平等,很难对地理位置分散的众多节点采用相同的安全措施.攻击者可以寻找安全薄弱的节点入侵区块链网络.

2) 区块链交易之间的关联性可以被用于推测敏感信息.区块链中所有交易都存储在公开的全局账本中,攻击者很容易获得所有交易信息.通过分析交易中的关联关系,攻击者能够逐步降低区块链地址的匿名性,甚至发现匿名地址对应用户的真实身份信息.

3) 区块链应用面临多种安全威胁.区块链技术仍处于发展初期,存在多种安全缺陷,有可能对区块链应用造成安全威胁.例如基于以太坊的众筹应用DAO因为黑客攻击,被窃取了300多万个以太币[15].

Table 1 The Advantages and Disadvantages in Aspect of the Privacy Protection on Blockchain Technology

表1 区块链技术隐私保护方面的优势和不足

Layer	Advantages	Disadvantages
Network	P2P network is difficult to monitor.	Nodes in blockchain network have a low performance and are vulnerable to being taken over.
Transaction	Blockchain technology supports anonymous transactions.	The correlation among transactions can be used to derive sensitive information.
Application	Decentralized application can effectively counter cyber attacks.	Blockchain applications face various security vulnerabilities.

2.2 网络层面临的隐私威胁

根据2.1节对网络层缺陷的分析可知,网络层的主要威胁是恶意节点可以轻易接入网络,监听网络层的通信数据.以比特币为例,能够获得信息包括:

1) 节点的IP地址.通过部署探针节点,攻击者可以搜集比特币网络中节点的IP地址.基于IP地址可以分析出网络规模、节点在不同国家的分布情况、节点的在线规律等,可以为进一步分析提供素材.我们利用开源程序搜集比特币网络中节点的IP信息.经过一周的搜集,发现比特币服务器节点IP 9 713个,客户端节点IP 36 514个.针对服务器节点,除了IP以外,还能搜到许多额外的身份信息.如图3所示.

IP	City	Country	AS	Organization
82.XX.XX.68	Baarn	Holland	AS9143	Ziggo
118.XX.XX.128	Hangzhou	China	AS37963	Hangzhou Alibaba Advertising Co., Ltd
122.XX.XX.20	Hangzhou	China	AS4134	Chinanet

Fig. 3 The server information of bitcoin node

图 3 比特币服务器节点信息

2) 节点之间的拓扑关系. 通过采用主动获取和被动监听的方式,探针节点可以用于搜集节点之间的拓扑关系. 比特币网络中节点的拓扑关系主要是指节点的邻居节点信息. 邻居节点最多包括 8 个外向节点和 117 个内向节点[16]. 基于节点拓扑关系,攻击者可以分析网络层信息的传播路径,确定信息的始发节点.

3) 网络传输信息. 比特币网络层传播的数据没有加密,攻击者可以直接读取网络中传播的交易信息的内容,例如发送地址和接收地址等. 通过将发送地址和信息的始发节点相关联,能够获得匿名地址背后的真实身份信息.

基于上述信息,攻击者有可能将网络层捕获的交易信息和始发节点的 IP 地址关联,从而对用户身份隐私造成威胁,这种分析方法被称为交易溯源技术.

区块链网络中,节点采用中继转发的模式传播信息,此模式已在 1.1 节介绍. 根据转发交易的时间顺序,可以将参与转发的节点分为始发节点和中继节点. 始发节点是指第 1 个发出信息的节点,中继节点是指除始发节点以外的节点. 如果能够找到网络中传播的交易信息对应的始发节点,就可以将交易中的输入地址和具体的 IP 信息关联,从而降低地址的匿名性. 但是,由于始发节点转发的交易数据和中继节点转发的交易数据在内容上并没有明显区别,很难分辨交易数据是否来自于始发节点. 因此,在网络层获取身份隐私的主要难点在于如何从众多的信息中找到由始发节点转发的交易数据.

Koshy 等人[17]通过分析比特币交易传播规律,发现可以利用特殊交易模式寻找始发节点. 比特币中交易信息采用中继转发的模式进行传播. 正常情况下每条交易信息将被多个节点转播,且每个节点只转发一次. 但是在一些特殊情况下,交易信息可能只被 1 个节点转发或者被个别节点多次转发. 这些特殊交易可以用于推测交易的始发节点. 例如只被 1 个节点转发的交易通常是由于交易格式存在问题,只被始发节点转发一次,其他节点由于验证失败拒绝转发. 利用特殊交易寻找始发节点的方法简单易行,但是由于特殊交易模式占所有交易的比例较小(论文实验中特殊交易的比例低于 9%),因此效果有局限.

Kaminsky[18]在 2011 年的黑帽大会上提出:"第一个告诉你交易的节点可能就是交易的始发节点". 攻击者只需要尽可能多地连接服务器节点,记录从不同节点转发的交易信息,然后可以直接判定首先转发信息到达探针节点的节点就是始发节点. 这种方法实现简单,推测的准确率将随着探针节点连接数的增加而不断提高. 但是这种方法存在一些缺陷:首先,比特币中的很多节点并不对外提供服务,例如部署在 NAT 服务后面的节点无法被探针节点连接. 其次,交易数据在网络中传播时面临网络延迟等干扰,始发节点转发的交易信息可能因为传播延迟落后于其他中继节点转发的信息.

Biryukov 等人[19]在 Kaminsky 提出的算法上进行优化,核心思路是始发节点转发交易后,始发节点的邻居节点会是第 2 批转发交易的节点,因此可以利用邻居节点转发信息的时间排序来定位始发节点. 作者将探针搜集的交易信息按照到达时间分类排序,然后针对每一类交易抽取排名前 8 的节点和 8 个邻居节点进行比对,如果重合率超过阈值,即认为节点是交易的始发节点. 这种方法扩展了判断依据,能够减少网络延迟等干扰条件的影响,有效提高推测的准确率. 实验显示识别准确率为 11%,如果采用一些辅助的攻击,准确率能提高到 60%. 此外,这种攻击方法可以发现隐藏在 NAT 服务后面的客户端节点.

2.3 交易层面临的隐私威胁

区块链技术中存储交易信息的全局账本是公开的,任何加入区块链网络的节点都可以获得完整的副本. 通过分析全局账本中的交易记录,潜在攻击者有可能对用户的交易隐私和身份隐私带来威胁.

1) 交易隐私威胁.攻击者可以通过分析交易记录获得有价值的信息.例如特定账户的资金余额和交易详情、特定资金的流向等.

2) 身份隐私威胁.攻击者在分析交易数据的基础上,可以通过结合一些背景知识获得交易者的身份信息.

我们以比特币为例,介绍交易层存在的隐私威胁.

2.3.1 交易隐私威胁

在比特币中,每一笔交易都是可追溯的.交易的输入地址来源于之前一笔交易的输出,交易的输出地址又会在其他交易中作为输入.根据交易之间的链式关系,分析人员可以获得2种信息:

1) 任何一笔资金的使用情况.比特币中的资金来源于"挖矿"过程产生的比特币,比特币被挖出后首先记录在矿工的挖矿地址上,然后通过交易的形式转移给其他地址.由于挖矿信息和交易信息都将记录在全局账本中.因此,通过分析这些公开数据,攻击者可以获得任何一笔比特币的所有交易过程.

2) 任何一个比特币地址的相关交易.每一笔交易中详细记录所有的输入地址和输出地址的信息.因此,分析人员可以在全局账本中通过检索特定地址发现全部相关交易.

基于上述信息,分析人员通过对具有关联性的区块链地址进行聚类分析,能够得到许多有价值的信息:

1) 发现不同地址之间的资金关系.账户之间的资金交易能够反映出很多有价值的信息,尤其对一些特殊账户,这点更为明显.Reid和Harrigan[20]针对维基解密公布的账户进行数据分析,能够统计出维基解密网站公布的比特币地址的资金余额、资金来源和资金流向.论文还针对比特币中公开的一个盗窃地址进行分析,发现与盗窃地址交易最密切的5个地址,揭示攻击者盗窃前的行为和盗窃后的资金流向.

2) 跟踪特殊交易.针对大额或者涉嫌盗窃等恶意行为的特殊交易,可以通过持续观察,监控特殊交易后续的交易信息,追踪资金流向.Liao等人[21]通过分析比特币交易数据,对勒索软件CryptoLocker的攻击行为进行了分析.CryptoLocker是一类勒索软件的统称,通过加密受害者的文件并以比特币的形式索取赎金.作者以勒索软件公开的比特币地址为起点,研究关联交易,最终找到了968个属于该组织的地址,鉴定出价值1 128.40 BTC的赎金交易.这些信息有助于发现犯罪分子的真实身份,遏制此类勒索事件.

3) 发现交易规律.研究数字货币的交易规律有助于提高系统的安全性.Ron和Shamir[22]关注比特币交易的统计数据.作者发现了364个单笔交易大于50 000 BTC的特殊交易,并针对其中一笔90 000 BTC的交易研究交易规律.作者发现这种大额交易会采用多种方式将资金分散到不同的账户,包括长链交易模式、分叉合并模式、循环模式、存储账户模式和二叉树模式.

2.3.2 身份隐私威胁

区块链交易中存在很多潜在的知识,利用这些潜在知识有可能推测出交易数据背后的身份隐私.

第1种潜在知识来源于比特币本身的设计:

1) 同一个交易中的所有输入地址都隶属于同一个用户集合(同一个人或者一个机构).由于多输入交易中的每个输入都需要单独签名,因此大多数多输入交易都是由同一个用户发起.这项推测条件被很多研究作为启发式推测条件使用[20-22],取得较好的效果.

2) 在同一个coinbase交易[23]中的多个输出地址属于同一个用户集合,例如矿池.这项推测条件的背景是很多矿工为了增加收入而加入矿池进行集体挖矿,挖矿完成后所有参与挖矿的矿工地址都会记录在coinbase交易的输出地址区域.

3) 找零地址和输入地址隶属于同一个用户.找零地址用于接收交易中的找零资金,并在以后的交易中由比特币程序自动选择作为输入地址.因此,如果能够识别出找零地址,就能够发现两个交易中输入地址之间的关联关系[24].找零地址的特征包括:作为输出地址的情况通常只会出现一次;找零地址不会同时出现在输入地址和输出地址;输出地址中不能只有找零地址.

此类的潜在知识是由于比特币本身的设计造成的,分析人员利用这些潜在知识,可以发现不同地址之间的关联性,降低区块链地址的匿名性.Meiklejohn等人[25]使用启发式的聚类分析技术分析区块链中的交易数据,能够识别出属于同一个用户的不同地址.作者针对丝绸之路网站的公开地址和一些盗窃案件相关的地址进行分析,发现了很多有关联的地址.Zhao[26]提出一种针对比特币交易数据的聚类过程,针对比特币全局账本中35 587 286个地址进行分析,得到13 062 822个不同的用户集合.

第2种潜在知识来源于用户使用区块链时产生的一些规律信息.例如:

1)交易特征.区块链全局账本中交易的特征通常和实际发生的交易过程相匹配.在日常生活中很多交易行为都有各自的特点,例如早餐店的交易经常发生在早晨,交易金额集中在1~20元左右;加油站交易时间比较平均,但是交易金额集中在几个特殊值100元、200元或者是加满的价格(随着油价的变化而变化,具有普遍的规律性).当用户使用区块链数字货币完成这些交易时,这种交易特征会反映在区块链交易中.因此,攻击者可以通过分析区块链交易数据背后的交易特征,给匿名账户进行身份画像,从而将匿名的区块链地址和用户真实身份匹配.Androulaki等人[27]在学校中设计模拟实验,学生使用比特币作为日常交易货币,分析人员采用基于行为的聚类技术对比特币交易数据进行分析.作者发现即使用户采用比特币推荐的隐私保护方法(一次性地址策略),也能够将40%的学生身份和区块链地址匹配.

2)交易规律.不同用户在使用区块链服务时有不同的行为规律.这些规律包括交易时间间隔(RTI)、资金流向(coinflow)、连通性(交易中输入地址和输出地址的数目)等.Monaco[28]对用户的交易规律进行抽象,提出12种参数(时间间隔 RTI、资金流向 CF、输入输出数量差 IOB 等),然后提出一种基于参数的身份识别方法.通过对6个月的比特币数据进行实验,作者发现最佳的参数组能够有效识别匿名区块链地址对应的真实用户身份,识别精度达到62%,同时错误率低于10.1%.

2.4 应用层面临的隐私威胁

应用层的主要角色包括使用区块链技术的用户和提供区块链服务的服务商.这两者在处理区块链业务时都有可能带来隐私泄露威胁.

1)用户行为导致的隐私泄露威胁.区块链是一种新兴技术,用户可能因为不了解区块链安全机制进行了一些可能泄露隐私信息的行为.例如,很多用户将自己的比特币地址发布在论坛或者其他社交网站上,攻击者可以通过社会工程学等方法将这些比特币地址和用户的真实身份关联.Meiklejohn等人[25]搜集了比特币论坛上公开的比特币地址,然后利用污点分析和聚类方法发现了很多比特币地址的身份信息.

2)区块链服务商导致的隐私泄露威胁.区块链技术本身提供了多种保护隐私的方法,但是很多提供区块链服务的网站存在显著的隐私泄露隐患.2014年,全球最大的比特币交易平台Mt.Gox遭遇大规模比特币盗窃案件,网站存储的涉及比特币的身份信息被泄露到黑客手中.2015年,比特币论坛BitcoinTalk遭受网络攻击,攻击者窃取了49.9万用户数据,包括用户名、电子邮箱、密码、生日、保密问题.

攻击者获得这些敏感信息后,能够将区块链全局账本中的匿名地址和真实用户相关联,掌握用户全部的交易信息,侵害用户的隐私.

3 区块链隐私保护机制

3.1 区块链隐私保护的特点

区块链技术采用一系列密码学算法在非信任节点之间建立信任关系,而不是依赖中心机构的信用背书,这种特殊的安全模型使得区块链下的隐私保护不同于传统的隐私保护,存在的差异主要分为2类:

1)隐私保护的侧重点不同.在传统隐私保护方案中,所有的隐私数据都被保存在可信节点的中心服务器内部,侧重点在于保护数据在存储阶段和传输阶段不被外泄.而在区块链技术中,为了维护分布式账本的一致性,保证交易的公信力,区块链中的所有交易数据必须公开给全网所有参与节点(注:部分区块链技术支持轻节点,不用存储全部数据,而是在需要时向其他全节点索要数据).因此,区块链技术无法采用中心存储的方式保护隐私,而是将侧重点落在保护交易的匿名性,即虽然所有的交易细节都是可见的,但是攻击者无法根据交易数据找到交易双方真实的身份信息,从而无法对用户造成损害.

2)隐私保护面临很多限制条件.区块链技术中运行区块链程序的节点通常是家用主机,而非传统服务器,很多复杂的隐私保护算法在区块链中是不实用的.此外,设计隐私保护算法时必须避免破坏区块链的共识机制.例如,对数据加密是保护隐私的常见方法,而在区块链中直接对交易信息加密,会使得其他节点无法验证交易的正确性,导致交易作废.

根据区块链技术的特点,我们将区块链中的隐私保护机制分为3类:

1)网络层的隐私保护.网络层包含底层通信的整个过程,包括区块链节点设置模式、节点通信机制、数据传输机制等.在区块链最早的应用比特币系统中,节点不需要审批,任何用户都可以通过运行比特币程序成为区块链的节点.节点之间利用P2P协议进行相互通信和数据传输.这种机制导致攻击者

不仅可以监听整个网络的通信信息,还可以主动和其他节点通信获取隐私数据.因此,网络层隐私保护的侧重点是限制节点的权利,对抗被动监听和主动攻击.

2) 交易层的隐私保护.交易层包含区块链中数据产生、验证、存储和使用的整个过程.区块链技术为了保证交易的可靠性、不可篡改性和分布式一致性,设计了特殊的数据结构和共识机制.这些机制保证了在分布式不可信的网络节点间维护统一的高公信力的账本.但是,这些机制也导致了隐私泄露风险.完整的、公开的交易账本不仅会泄露交易数据,还会泄露数据背后的交易者之间的关系,甚至泄露身份隐私.因此,交易层隐私保护的侧重点是满足区块链基本共识机制的条件下,尽可能隐藏数据信息和数据背后的知识.

3) 应用层的隐私保护.应用层包含区块链技术被外部使用的过程.外部使用即包括普通用户使用区块链程序,也包括其他应用程序调用区块链的接口.区块链被外部使用的过程存在泄露交易隐私和身份隐私的威胁.例如,用户在论坛等社交网站公布自己的比特币地址.因此,应用层隐私保护的侧重点包括提升用户的安全意识、提高区块链服务商的安全防护水平.

3.2 网络层的隐私保护机制

通过分析网络层的攻击方法,可以看出攻击者主要是通过监听网络层信息来搜集交易隐私和身份隐私.因此,网络层防御机制的重点是增加攻击者搜集网络层数据的难度.

现有的防御机制可以分为3类:

1) 限制接入.对区块链中的节点进行授权控制,没有得到授权的节点无法接入网络,不能获得交易信息和区块信息,这将从根本上增加网络层攻击的难度.但是,这种方法需要修改区块链的运行机制,目前主要运用在私有链或者联盟链的架构中.例如,在超级账本(hyperledger)[29]中,所有节点必须经过CA节点的认证才能接入网络.

2) 恶意节点检测和屏蔽.在公有链架构中,不能直接限制节点接入网络,但是可以采取检测机制,发现恶意节点并加入黑名单,阻止恶意节点继续搜集敏感信息.Huang等人[30]提出一种基于行为模式聚类的恶意节点检测方法,能够快速定位恶意节点,消除恶意节点带来的隐私泄露隐患.

3) 网络层数据混淆.为了阻止攻击者通过发现网络拓扑获得身份隐私信息,一些研究人员提出可以将区块链运行在具有隐私保护特性的网络上,例如洋葱网络(Tor)[31].洋葱网络是一种应用层的匿名通信技术,通信数据首先被多层加密然后再由若干个被称为洋葱路由器组成的通信线路上传送,攻击者很难发现发送者的真实IP.目前洋葱网络是比特币官方推荐的保护隐私的方法.另外一种以隐私保护著称的数字货币门罗币采用一种替代Tor的匿名通信协议I2P[32].相对于Tor协议使用同一条网络链路实现数据的发送和接收,I2P使用多条链路发送数据和接受数据,能够更好地隐藏IP,防止通过网络层信息实现交易溯源[33].

3.3 交易层的隐私保护机制

通过分析交易层的攻击方法,可以看出攻击者主要是通过分析公开的区块链交易数据获得隐私信息.因此,交易层保护机制的侧重点是在满足区块链正常运行的基础上,防止恶意节点获得准确的交易数据.目前,研究人员已经提出多种交易层的隐私保护方案,这些方案在技术架构、实现方法、实现策略上各不相同,能够满足不同的隐私保护需求.区块链技术从数据存储的角度可以看做是一种分布式数据库,因此我们将不同的保护机制按照数据库隐私保护的分类方法进行3种分类:

1) 基于数据失真的技术.通过将交易内容的部分数据进行混淆,使攻击者无法获得准确的数据,增加分析难度.这种方案的难点在于不破坏交易结果的条件下,防止攻击者发现不同地址之间的交易关系.

2) 基于数据加密的技术.通过将交易信息加密,使攻击者无法获得具体的交易信息,从而无法开展分析.这种方案的难点在于实现加密的同时,必须保证原有的验证机制不受影响.

3) 基于限制发布的技术.通过发布少量或者不发布交易数据,减少攻击者能够获得的交易数据,增加分析难度.

3.3.1 基于数据失真的保护方案

由于区块链交易会详细记录在全局账本中,攻击者可以通过分析交易内容发现输入地址和输出地址之间的关系,进而推测出交易隐私和身份隐私.为了对抗这种攻击,一种直观的方法是在不改变交易结果的前提下对交易内容进行混淆,增加攻击者的分析难度.这种防御方法在数字货币领域应用广泛,被称为"混币"机制.混币机制最早来源于Chaum 1981年发表的文章[34],最早是用于实现双方之间的匿名通信.基本思想可表达为

$$C_M(Z_1,C_A(Z_0,m),A)\rightarrow C_A(Z_0,m),A, \quad (1)$$

式(1)左侧为发送方发给中间人的信息,式(1)右侧为接收者从中间人获取的信息.发送方使用中间人的公钥 C_M 对内容加密,然后发送给中间人.在发送过程中,即使信息被攻击者截获,也不能解密.中间人收到信息后使用自己的私钥对内容解密,能够得到"Z_1,$C_A(Z_0,m)$,A".其中,"Z_1"用于确保传送信息的准确性,中间人验证合法后将丢弃此参数."$C_A(Z_0,m)$,A"是用接收者的秘钥"C_A"进行加密的内容,中间人无法解密,A 是通信目的地址.中间人将把密文"$C_A(Z_0,m)$"发送给地址 A.接收方使用自己的秘钥解密"$C_A(Z_0,m)$",就可得到传送的密文内容"Z_0"和"m".

通过采用中间人中转信息,攻击者无法准确判定发送者和接收者是否进行通信.此方法还可以通过采用多个中间人中转的方式提升攻击者的分析难度.

数字货币中的混币机制借鉴了上述思想,即采取机制隐藏交易中输入地址和输出地址之间的关系.由于这种机制中交易数据没有减少,只是对交易来源和交易去向进行模糊,因此可以归类于数据失真的隐私保护技术.

在数字货币领域,混币技术主要可以分为基于中心节点的混币方法和去中心化的混币方法.

1) 基于中心节点的混币方法.此类方案的核心特点是混币过程由第三方节点执行.参与混币的用户首先将资金发送给第三方节点,然后第三方节点对资金进行多次交易,最终将资金转移给参与用户指定的地址.由于资金经过第三方节点的处理,攻击者很难发现参与混币用户的资金流向.

此类方法简单易行,不需要额外的技术改进,适用于比特币以及其他数字货币.目前有很多网站提供这种混币服务.例如 Bitlaunder[35],Bitcoin Fog[36],Blockchain.info[37].用户通过支付混币费用,就可以使用网站提供的混币服务.

但是,这种方法由于需要第三方节点提供混币服务,存在很多天然的缺陷,包括:

① 额外的收费和较慢的混币速度.提供混币服务的节点通常会收取混币费用,而且随着混币次数的增加,费用会直线上升.此外,混币的时间也会随之增加.通常的延迟时间为 48 h,交易费用为 1%~3%.

② 存在盗窃资金的风险.此方案中,第三方节点收到用户的资金后有可能不履行协议,盗窃用户的资金.用户没有有效的反制措施.

③ 中间节点可能泄露混币过程.此方案中第三方节点了解全部的混币过程,用户无法保证第三方节点不会泄露混币过程信息.

针对这些缺陷,出现了很多改进的方法.Bonneau 等人[38]提出一种改进的中心化混币方案 mixcoin.这种方案增加了审计功能,一旦第三方节点违规操作,用户可以公布签名数据,使违规的中间人迅速丢掉声誉,不能继续提供混币服务.Valenta 和 Rowan 在 mixcoin 的基础上,采用盲签名技术对中心化混币方案进行进一步优化,他们设计的 Blindcoin 方案[39]能够保证第三方节点正常提供混币服务的同时,不能建立起输入地址和输出地址的映射关系,因此能够防止第三方泄露混币过程信息.ShenTu 和 Yu[40]提出一种基于椭圆曲线的盲签名混币方案,能够在保证匿名性的基础上提升计算效率.2015 年上线运营的匿名数字货币达世币(DASH)[41],从经济学的角度解决中心化混币方案面临的威胁.达世币中执行混币过程的中心节点被称为主节点,所有主节点必须向系统支付 1 000 达世币(达世币中的数字货币)的押金才能获得执行混币操作的权利.通过设置押金,增加了主节点违规操作的代价.

2) 去中心化的混币方法.此类方案的核心特点是混币过程不需要第三方节点执行.最早的方案是由 Gregory Maxwell 在比特币论坛上提出的 CoinJoin 机制[42],核心思想是通过将多个交易合并成 1 个交易,隐藏交易输入方和输出方的对应关系.对于一个多输入-多输出交易,潜在攻击者无法通过阅读交易信息有效区分输入和输出之间的关系.CoinJoin 思想被运用在多种匿名比特币交易中,例如 Dark Wallet[43],CoinShuffle[44] 和 JoinMarket[45].

CoinJoin 机制能够增强所有用户的隐私保护能力.一旦数字货币系统中部分节点采用 CoinJoin 协议,即使其余用户没有使用这种协议,也不能采用原有的推测方法,认为一个交易中的多个输入地址隶属于同一个用户.CoinJoin 方案不依赖第三方节点,能够有效避免中心化混币方案存在的资金偷窃、混币费用等问题.但是由于没有中心节点,CoinJoin 方案中参与混币的用户必须自行协商和执行混币过程,这存在许多缺陷:

① 在寻找参与混币用户的过程中,可能需要中心节点,面临中心化混币同样的威胁.

② 在节点协商的过程中,参与混币的节点可能会发现其他节点的混币信息.

③ 在执行混币过程中,如果部分节点违规操作,可能导致混币过程失败.攻击者可以利用这点以低成本实现拒绝服务攻击.

④ CoinJoin 方案形成的多输入多输出交易将记录在全局账本中,用户无法抵赖他们曾经参与过混币.

针对这些缺陷,出现了很多改进的方法. Ruffing 等人提出一种完全去中心化的比特币混币协议 CoinShuffle[46]. CoinShuffle 方案在 CoinJoin 的基础上设计一种输出地址洗牌机制,能够在不需要第三方的条件下完成混币过程,还能保证混币参与方不知道其他交易方的对应关系. 但是 CoinShuffle 方案在混币过程中要求参与者同时在线,容易遭受拒绝服务攻击. Bissias 等人设计一种能够利用区块链中的广告信息匿名发现混币参与方的去中心化混币协议 Xim[47]. Xim 采用一种多轮两方混币协议,具有可调控的成功率. 与 CoinJoin 机制相比,Xim 方案中恶意节点发动攻击的代价将随着参与混币用户的数量线性增加,能够有效对抗女巫攻击和其他拒绝服务攻击. CoinParty[48] 方案采用安全多方计算协议实现了一种改进方案,能够在部分混合节点恶意操作或者失效的情况下,保证混币过程的有效性. 门罗币(Monero)[49] 是一种以隐私保护为主要特征的新型数字货币,采用环签名机制实现混币过程. 相对于其他方案,门罗币中用户实施混币过程时不需要和其他用户交流. 任何一个用户可以自行实现混币,能够有效杜绝去中心化混币方案面临的拒绝服务攻击、混币参与用户泄露混币过程等问题.

混币方案操作简单、适用性广,在区块链数字货币中应用广泛,有很多改进方案. 我们从是否依赖第三方、是否存在盗窃风险、是否需要混币费用等多个方面进行对比分析. 如表 2 所示:

Table 2 Comparison of Mix Mechanism in Blockchain

表 2 混币机制特征对比

Protocol	Rely on Third Party	Mix Cost	Risk of Theft	Risk of DOS
Mix	√	√	High	Low
Mixcoin	√	√	Middle, Support Auditing	Low
BlindCoin	√	√	Middle, Blind Signature	Low
DASH	√	√	Middle, Deposit	Low
CoinJoin	×	×	Low	High
CoinShuffle	×	×	Low	High
Xim	×	√	Low	Low
CoinParty	×	×	Low	Low
Monero	×	×	Low	Low

3.3.2 基于加密机制的保护方案

加密机制是隐私保护领域的常用方案. 通过对敏感数据加密,确保只有持有秘钥的用户能够阅读数据,其他人即使获得密文也无法解密,从而避免数据泄露. 传统区块链应用中数据是明文存储的,任何节点都可以维护数据副本,验证新产生的交易和区块,这是区块链交易公信力和可靠性的基础. 因此,在区块链中采用加密技术保护隐私必须保证节点可以在加密数据上完成交易验证任务. 此外,由于区块链交易需要由所有节点共同验证,因此必须减少加密机制对验证效率的影响.

区块链中需要加密的对象主要是具体的交易信息,包括交易的来源、去向和交易的内容. 在数字货币应用中,已经出现一些基于加密的保护方案:

1) 门罗币的加密方案. 门罗币是一种专注于隐私保护的数字货币,采用多种隐私保护机制[50]. 其中,为了防止外部人员从账本中发现一笔资金的真实去向,门罗币对交易输出地址进行加密. 传统的数字货币中,交易输出地址的内容是接收方的公钥信息和地址信息,观察者可以直接发现资金的去向. 在门罗币中,输出地址是由接收方的公钥和发送方产生的随机参数加密后得到的新地址信息,由于随机参数只有发送方掌握,因此观察者无法发现新地址信息和接收方之间的关系. 通过产生不同的随机参数,可以保证每一次交易的输出地址都不同,而且相互之间没有关联关系.

2) Zcash 的加密方案. Zcash 是一种新型的数字货币[51],前身为 Zerocoin[52] 项目. Zcash 使用承诺函数将每一笔交易的来源、去向和金额封装到若干参数中,同时使用零知识证明技术 zk-SNARKs[53] 来证明交易. 证明过程不需透露相关信息,因而可以隐藏区块链交易的发送方、收款方乃至交易的价值. 花费 Zcash 需要的若干参数由交易的发起方使用非对称加密技术进行加密,只有掌握正确查看密钥的用户才可访问这些内容. Zcash 是目前隐私保护最好的数字货币,但其采用 zk-SNARKs 算法生成证明的过程非常缓慢,通常需要 1 min 才能生成新的证明,在效率上存在瓶颈.

3.3.3 基于限制发布的保护方案

限制发布方案是指直接将涉及隐私的数据从公开数据库中移除. 相比混币机制和加密机制,这种思路直接彻底,能够从根本上保证隐私数据的安全. 但是,这种方法对业务场景的限制较多,需要对协议底层进行较多修改. 常见方案包括 2 种:

1) 闪电网络.闪电网络是比特币中的一种微支付技术[54],用于提供可靠的链外交易(交易细节不需要记录在区块链上),从而提升区块链数字货币的交易规模,满足微支付需求.闪电网络技术实施后,用户之间的大部分交易细节在线下执行,只有第一次交易和最后一次交易需要记录在区块链账本上,因此能够有效保护交易隐私.

2) 联盟链和私有链.传统的区块链应用大多数是基于公有链的,例如比特币、以太币.在公有链应用中,任何人都可以自由加入区块链网络,维护全部的交易数据,这使得公有链应用具有很高的公信力,但是这也带来了身份隐私和数据隐私的威胁.为了更好地保护隐私,区块链技术产生了联盟链和私有链的分支.联盟链中,多个行业单位构成联盟,只有联盟内的成员能够维护区块链数据,其他非授权节点不能接触区块链数据;私有链中,只有内部用户才能维护区块链数据.这 2 种新的架构从根本上关闭了非授权节点接触数据的渠道,显著降低隐私泄露的风险.

3.4 应用层的隐私保护机制

通过分析应用层的攻击方法,可知看出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.因此应用层防御机制的重点是从用户的角度提升保护能力.

用户可以采用的防御方法通常有 2 种:

1) 使用具有隐私保护机制的区块链应用.比特币是区块链技术在数字货币领域的第 1 个应用,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(Zcash).

达世币在比特币的基础上增加了基于主节点的混币策略,能够隐藏资金流向.由于主节点需要支付大量押金,同时用户可以配置使用多个主节点进行混币操作,达世币的混币过程能够有效减少同类混币服务面临的盗窃资金、泄露混币过程等问题.

门罗币采用环签名技术模糊交易的输入地址,增加攻击者分析资金来源的难度.由于环签名过程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.

Zcash 是目前隐私保护效果最好的数字货币.通过采用 zk-SNARKs(简洁的非互动性零知识证明)技术,能够在满足验证和共识机制的条件下隐藏区块链交易的发送方、收款方乃至交易的金额.

这些新型的数字货币采用密码学技术保护交易数据,相对比特币能够更好地保护用户的身份隐私和交易隐私.

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.以比特币为例,冷钱包通过将秘钥离线保存,能够有效防止黑客攻击,但是有可能出现存储介质丢失和被盗带来的安全风险,隐私保护的关键是保护存储介质的安全性,可以采用多重备份、加密存储等机制保护存储介质的安全.本地钱包面临黑客窃取钱包文件的风险,可以采用钱包加密、修改默认存储位置、变换文件名等方式保护钱包文件.在线钱包的安全威胁主要是钱包服务器被黑客攻击,在比特币历史上已经出现了多起严重的数据泄露事件.目前在线钱包网站主要采用冷钱包[55]和多重签名技术保护账户隐私.表 3 介绍各种钱包程序的特点.

Table 3 Comparison of Blockchain Wallet

表 3 钱包程序对比

Category	Secret Key Saving Mode	Major Threats	Protection Mechanism
Cold Wallet	Offline device, paper wallet, brain wallet	Device lost or damaged	redundant backup, encrypted storage
Local Wallet	Save on the host or mobile device	cyber attack or device failure	Encrypt wallet, wallet backup
Online Wallet	Save in server	server was attacked	Multiple signature

4 未来研究方向

区块链技术发展日新月异,隐私保护的重要性持续提升.但是,目前的隐私保护方案都存在一些缺陷,需要继续进行研究.

在网络层,虽然现有的交易溯源技术准确率低,还不具备大规模实施的条件,但是网络层的安全威胁在区块链技术体系中具有通用性,凡是采用 P2P 协议作为底层通讯协议的区块链应用都存在这个隐患.现有的保护方案中,限制接入的方案只适用于联盟链和私有链,在公有链中很难实施;恶意节点检测

和屏蔽机制是事后的补救措施,而且必须配合限制接入机制才能起到较好的防御效果.否则,攻击者可以随时通过更换 IP 等方式重新部署恶意节点;采用匿名通信协议能够增加攻击者监听网络层信息的难度.但是现有的匿名通信协议在效率、性能、易用性上还存在缺陷.例如比特币官方推荐采用洋葱网络(Tor)保护隐私,但是这种方案已经被发现存在漏洞,攻击者可以利用 DOS 攻击强迫节点退出正常的 Tor 节点,然后再进行攻击[56].此外,匿名通信协议通常伴随着过于繁琐的使用方法,这甚至会使部分用户更容易遭受恶意软件的威胁.

在交易层,基于数据失真的隐私保护方案实现简单,适用于比特币等现有的区块链应用.但是此类方法的保护效果有限,分析人员仍有可能通过分析交易之间的关联性发现隐藏的信息;基于加密机制的保护方案能够有效应对基于数据分析的攻击,但是此类方法必须对底层协议进行大幅改动,不适用于比特币等现有的区块链应用.此外,此类方案通常需要耗费更多的计算资源,在效率上存在瓶颈;基于限制发布的技术能够有效降低敏感数据的数量,从源头上降低隐私泄露的可能性.但是此类技术通常需要修改底层协议,实施难度较大.此外,这种技术由于改变了传统区块链技术中所有交易数据公开的特点,需要解决交易可靠性和共识机制面临的问题.

在应用层,现有的方案主要是通过提高用户安全意识和能力来减少隐私泄露的风险.然而,当用户数量增加时,很难保证所有用户具有相同的安全意识和防护能力.

基于上述问题,我们提出 3 项未来的研究方向:

1) 按需配置的网络层安全防护机制.针对联盟链和私有链,采用合适的访问控制策略防止恶意节点接入和监听网络,从根本上增强网络层的保护能力.此外,联盟链或者私有链与传统中心化架构有很多相似之处,可以采用传统中心化架构中成熟的安全措施.针对公有链网络,重点研究异常节点检测的方法,及早发现和屏蔽恶意节点.此外,需要研究在效率、性能、易用性方面更好的匿名通信机制,替代现有的 Tor 等匿名通信方案.

2) 基于密码学算法的交易层隐私保护机制.随着数据分析技术的发展,传统的混币机制保护隐私的效果将逐渐降低.有必要研究采用密码学算法保证混币的安全性,例如零知识证明机制和同态加密机制.基于加密的保护方案应该充分考虑区块链服务器在计算性能和存储性能上的缺陷,设计通用性更高的加密方法.

3) 安全秘钥技术.在应用层,除了提升用户安全意识、增强区块链服务商安全能力以外,重点是要研究钱包的密钥保护技术,开发使用方便、安全可靠的钱包程序.钱包秘钥直接关系到账户安全,可以研究基于口令、硬件以及生物特征等多因素认证机制,增强私钥的安全性.

此外,在研究区块链隐私保护技术的同时,也应该关注如何对滥用区块链技术的非法行为进行监管.目前使用区块链技术进行洗钱、勒索以及其他犯罪活动的事件层出不穷,例如影响全球 30 万名用户的勒索病毒 wannacry 就是使用比特币勒索赎金.由于比特币去中心化、匿名化等特征,很难阻止勒索行为,追踪勒索者的身份信息.

针对这种监管需求,主要有 2 类解决方案:

1) 加强行政监管.锁定区块链技术与现实社会的集合点,例如交易所等区块链服务商.目前大部分国家都出台规定,要求数字货币交易所等区块链服务机构实施 KYC 政策(即充分了解你的客户).通过登记用户身份以及检查大额数字货币交易,能够有效控制利用区块链技术实施非法活动的规模.

2) 加快监管技术研究.区块链技术的去中心化架构使得很难从根本上禁止区块链应用,很多区块链应用可以不依赖于外部服务商独立运行.例如,比特币系统中的用户可以不依赖于交易所直接进行数字货币交易,或者在境外交易所交易.此外,各种混币技术增加了区块链交易的监管难度.因此,除了行政手段外,有必要研究针对性的监管技术,检查和遏制利用区块链技术进行的非法活动.目前已经出现了很多专门从事区块链监管科技的公司和研究机构.美国纽约的公司 Chainalysis 开发了用于打击网络犯罪活动的工具,已经检查了价值 150 亿美金的比特币交易[57].美国桑迪亚国家实验室受美国政府支持开发分析工具,这种工具将帮助执法部门将比特币交易去匿名化[58].英国伦敦的区块链情报公司 Elliptic 为全球企业和执法机关提供数字货币监控支持,公司在 2016 年 3 月份收到 500 万美元有政府背景的投资[59].加拿大公司 Blockchain Intelligence Group(BIG)开发了 QLUE 来帮助世界各地的执法机构通过识别和追踪比特币来打击涉及比特币的金融犯罪交易[60].

5 总 结

本文介绍了区块链技术中隐私保护遇到的问题和挑战.首先从身份隐私和交易隐私的角度定义了区块链中隐私的概念;其次分别从网络层、交易层和应用层详细阐述了区块链隐私保护面临的威胁及其防护对策;最后针对区块链隐私保护的威胁和研究现状,展望了未来可能的研究方向.

参 考 文 献

[1] 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)

[2] Chris B, Adam W. Bitcoin ringing the bell for a new asset class [EB/OL]. [2017-06-10]. http://research.ark-invest.com/bitcoin-asset-class

[3] Gartner. Top 10 strategic technology trends for 2017 [EB/OL]. [2017-06-10]. http://www.gartner.com/technology/topics/trends.jsp

[4] Wang Jiye, Gao Lingchao, Dong Aiqiang, et al. Block chain based data security sharing network architecture research [J]. Journal of Computer Research and Development, 2017, 54(4): 742-749 (in Chinese)
(王继业, 高灵超, 董爱强, 等. 基于区块链的数据安全共享网络体系研究[J]. 计算机研究与发展, 2017, 54(4): 742-749)

[5] Au M H, Liu J K, Fang Junbin, et al. A new payment system for enhancing location privacy of electric vehicles [J]. IEEE Trans on Vehicular Technology, 2014, 63(1): 3-18

[6] Mihaylov M, Jurado S, Avellana N, et al. NRGcoin: Virtual currency for trading of renewable energy in smart grids [C] // Proc of the 11th Int Conf on the European Energy Market. Piscataway, NJ: IEEE, 2014: 1-6

[7] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system [EB/OL]. [2017-08-01]. http://www.bitcoin.org/bitcoin.pdf

[8] Bitnodes. Global bitcoin nodes distribution [EB/OL]. [2017-06-10]. https://bitnodes.21.co/

[9] Shawn W, Tome B, Josh B, et al. Storj: A peer-to-peer cloud storage network [EB/OL]. [2017-06-10]. https://storj.io/storj.pdf

[10] Dwork C, Naor M. Pricing via processing or combatting Junk Mail [C] // Proc of the 12th Annual Int Cryptology Conf Proceedings. Piscataway, NJ: IEEE, 1992: 139-147

[11] Larimer D. Transactions as proof-of-stake [EB/OL]. [2017-06-10]. https://bravenewcoin.com/assets/Uploads/TransactionsAsProofOfStake10.pdf

[12] Castro M, Liskov B. Practical byzantine fault tolerance and proactive recovery [J]. ACM Trans on Computer Systems, 2002, 20(4): 398-461

[13] Zhou Shuigeng, Li Feng, Tao Yufei, et al. Privacy preservation in database applications: A Survey [J]. Chinese Journal of Computers, 2009, 32(5): 847-861 (in Chinese)
(周水庚, 李丰, 陶宇飞, 等. 面向数据库应用的隐私保护研究综述[J]. 计算机学报, 2009, 32(5): 847-861)

[14] Antonopoulos A M. Mastering Bitcoin [EB/OL]. [2017-06-10]. https://www.bitcoinbook.info/

[15] Andy D. THE DAO [EB/OL]. [2017-06-10]. http://ethfans.org/posts/127

[16] Bonneau J, Miller A, Clark J, et al. SoK: Research perspectives and challenges for Bitcoin and cryptocurrencies [C] //Proc of the 2015 IEEE Symp on Security and Privacy (SP). Piscataway, NJ: IEEE, 2015: 104-121

[17] Koshy P, Koshy D, Mcdaniel P. An analysis of anonymity in bitcoin using P2P network traffic [G] //Financial Cryptography and Data Security. Berlin: Springer, 2014: 469-485

[18] Kaminsky D. Black Ops of TCP/IP 2011 [EB/OL]. [2017-08-01]. https://dankaminsky.com/2011/08/05/bo2k11/

[19] Biryukov A, Khovratovich D, Pustogarov I. Deanonymisation of clients in Bitcoin P2P network [C] //Proc of the 21st ACM Conf on Computer and Communications Security. New York: ACM, 2014: 15-29

[20] Reid F, Harrigan M. An analysis of anonymity in the bitcoin system [C] //Proc of the 3rd IEEE Int Conf on Privacy, Security, Risk and Trust. Piscataway, NJ: IEEE, 2011: 1318-1326

[21] Liao K, Zhao Ziming, Doupe A, et al. Behind closed doors: Measurement and analysis of cryptoLocker ransoms in bitcoin [C] //Proc of the 2016 APWG Symp on Electronic Crime Research (eCrime). Piscataway, NJ: IEEE, 2016: 1-13

[22] Ron D, Shamir A. Quantitative analysis of the full bitcoin transaction graph [G] //Financial Cryptography and Data Security. Berlin: Springer, 2013: 6-24

[23] Bitcoinwiki. Coinbase [EB/OL]. [2017-06-10]. https://en.bitcoin.it/wiki/Coinbase

[24] Bitcoinwiki. Change [EB/OL]. [2017-06-10]. https://en.bitcoin.it/wiki/Change

[25] Meiklejohn S, Pomarole M, Jordan G, et al. A fistful of bitcoins: Characterizing payments among men with no names [C] //Proc of the 13th ACM Internet Measurement Conf. New York: ACM, 2013: 127-140

[26] Zhao Chen. Graph-based forensic investigation of bitcoin transactions [D]. Ames, Iowa: Iowa State University, 2014

[27] Androulaki E, Karame G O, Roeschlin M, et al. Evaluating user privacy in bitcoin [C] //Proc of the 17th Int Conf on Financial Cryptography and Data Security. Okinawa, Japan: Financial Cryptography, 2013: 34-51

[28] Monaco J V. Identifying bitcoin users by transaction behavior [C] //Proc of 2015 SPIE DSS. Baltimore, Maryland: SPIE, 2015

[29] Hyperledger. Hyperledger architecture working group paper [EB/OL]. [2017-06-10]. https://www.hyperledger.org/

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

[31] Tor. Getting up to speed on Tor's past, present, and future Tor [EB/OL]. [2017-06-10]. http://www.theonionrouter.com/docs/documentation.html.en

[32] I2P. What does I2P do for you? [EB/OL]. [2017-06-10]. https://geti2p.net/en/

[33] Monero. What is Monero? [EB/OL]. [2017-06-10]. https://getmonero.org/get-started/what-is-monero/

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

[35] BitLaunder. BitLaunder's mixer vs "major exchanges" mixer [EB/OL]. [2017-06-10]. https://bitcoin.stackexchange.com/questions/25722/bitlaunders-mixer-vs-major-exchanges-mixer/25753

[36] Bitcoin Fog. Accessing bitcoin fog [EB/OL]. [2017-06-10]. http://bitcoinfog.info/

[37] Blockchain. Wallet [EB/OL]. [2017-06-10]. https://Blockchain.info/wallet/

[38] Bonneau J, Narayanan A, Miller A, et al. Mixcoin: Anonymity for bitcoin with accountable mixes [C] //Proc of the 18th Int Conf on Financial Cryptography and Data Security Financial. Barbados: Financial Cryptography, 2014: 486-504

[39] Valenta L, Rowan B. Blindcoin: Blinded, Accountable Mixes for Bitcoin [G] //Financial Cryptography and Data Security. Berlin: Springer, 2015: 112-126

[40] Shentu Qingchun, Yu Jianping. A blind-mixing scheme for bitcoin based on an elliptic curve cryptography blind digital signature algorithm [EB/OL]. [2017-06-10]. https://arxiv.org/ftp/arxiv/papers/1510/1510.05833.pdf

[41] Dash. Dash is digital cash [EB/OL]. [2017-06-10]. https://www.dash.org/

[42] Gregory M. CoinJoin: Bitcoin privacy for the real world [EB/OL]. [2017-06-10]. http://bitcointalk.org/index.php?topic=279249.0

[43] Andy G. Dark wallet' is about to make bitcoin money laundering easier than ever [EB/OL]. [2017-06-10]. https://www.wired.com/2014/04/dark-wallet/

[44] Kyle T. CoinShuffle aims to improve privacy in bitcoin [EB/OL]. [2017-06-10]. http://insidebitcoins.com/news/coinshuffle-aims-to-improve-privacy-in-bitcoin/29269

[45] Belcher. Joinmarket-Coinjoin that people will actually use [EB/OL]. [2017-06-10]. http://bitcointalk.org/index.php?topic=919116.0

[46] Ruffing T, Moreno-Sanchez P, Kate A. CoinShuffle: Practical decentralized coin mixing for bitcoin [G] //Computer Security (ESORICS 2014). Berlin: Springer, 2014: 345-364

[47] Bissias G, Ozisik A P, Levine B N, et al. Sybil-resistant mixing for bitcoin [C] //Proc of the 2015 ACM Workshop on Privacy in the Electronic Society. New York: ACM, 2014: 149-158

[48] Ziegeldorf J H, Grossmann F, Henze M, et al. CoinParty: Secure multi-party mixing of bitcoins [C] //Proc of the 5th ACM Conf on Data and Application Security and Privacy. New York: ACM, 2015: 75-86

[49] Monero. About monero [EB/OL]. [2017-06-10]. https://getmonero.org/knowledge-base/about

[50] Monero. A note on chain reactions in traceability in cryptoNote 2.0 [EB/OL]. [2017-06-10]. https://lab.getmonero.org/pubs/MRL-0001.pdf

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

[52] Miers I, Garman C, Green M, et al. Zerocoin: Anonymous distributed E-Cash from bitcoin [C] //Proc of the 2013 IEEE Symp on Security and Privacy (SP) Conf. Piscataway, NJ: IEEE, 2013: 397-411

[53] Ben-Sasson E, Chiesa A, Genkin D, et al. SNARKs for C: Verifying program executions succinctly and in zero knowledge [G] //Advances in Cryptology (CRYPTO 2013). Berlin: Springer, 2013: 90-108

[54] Joseph P, Thaddeus D. The bitcoin lightning network: Scalable Off-Chain instant payments [EB/OL]. [2017-06-10]. http://lightning.network/lightning-network-paper.pdf

[55] Okcoin. OKCoin cold wallet security design and protocol [EB/OL]. [2017-06-10]. https://www.okcoin.com/security.html

[56] Biryukov A, Pustogarov I. Bitcoin over Tor isn't a Good Idea [C] //Proc of the 2015 IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2014: 122-134

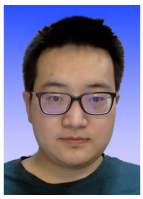
[57] Chainalysis. Protecting the integrity of digital assets [EB/OL]. [2017-06-10]. https://www.chainalysis.com/

[58] Sandia. Beating bitcoin bad guys [EB/OL]. [2017-06-10]. http://www.sandia.gov/news/publications/labnews/articles/2016/19-08/bitcoin.html

[59] Elliptic. The global standard for blockchain intelligence [EB/OL]. [2017-06-10]. https://www.elliptic.co/

[60] Blockchaingroup. Blockchain intelligence group [EB/OL]. [2017-06-10]. https://Blockchaingroup.io/

Zhu Liehuang, born in 1976. PhD, professor. His main research interests include cryptography, network and information security.



Gao Feng, born in 1987. PhD candidate. His main research interests include blockchain, cloud computing security and data privacy.

Zheng Baokun, born in 1978. PhD candidate. His main research interests include blockchain, network and information security.

Shen Meng, born in 1988. PhD, assistant professor. His main research interests include network security and privacy-preserving algorithms in cloud computing.

Mao Hongliang, born in 1990. PhD. Engineer of National Computer Network Emergency Response Technical Team/Coordination Center of China. His main research interest is blockchain application.

Li Yandong, born in 1991. MS candidate. His main research interests include blockchain, cloud computing security and data privacy.

Wu Zhen, born in 1976. PhD. Senior engineer of National Computer Network Emergency Response Technical Team/Coordination Center of China. His main research interests include cyber security and blockchain application.